Seria *Collegium. Psihologie* este coordonată de Adrian Neculau.

Dennis Howitt este lector în domeniul psihologiei sociale la Departamentul de Ştiinţe Sociale al Loughborough University. Dintre lucrările sale amintim: *First Steps in Research and Statistics: A Practical Workbook for Psychology Students* (în colaborare, Routledge, 2000), *The Sage Dictionary of Statistics* (în colaborare, Sage Publications, 2004) şi *Introduction to Research Methods in Psychology* (în colaborare, Prentice Hall, 2004).

Duncan Cramer este conferenţiar în domeniul psihologiei sănătăţii la Departamentul de Ştiinţe Sociale al Loughborough University. Între zonele sale de interes se numără sănătatea mentală, psihologia personalităţii şi a relaţiilor interpersonale, psihoterapia şi consilierea. Dintre lucrările sale amintim: *An Introduction to Statistics in Psychology: A Complete Guide for Students* (în colaborare, Prentice Hall, 2000) şi *Quantitative Data Analysis with SPSS 12 and 13: A Guide for Social Scientists* (în colaborare, Routledge, 2005).

Pentru informaţii suplimentare, vizitaţi ghidul web al volumului *Introducere în SPSS pentru psihologie* la adresa www.pearsoned.co.uk/howitt.

www.polirom.ro

Editura POLIROM
Iaşi, B-dul Carol I nr. 4, P.O. BOX 266, 700506
Bucureşti, B-dul I.C. Brătianu nr. 6, et. 7, ap. 33, O.P. 37, P.O. BOX 1-728, 030174

Descrierea CIP a Bibliotecii Naţionale a României

HOWITT, DENNIS
Introducere în SPSS pentru psihologie : versiunile SPSS 10, 11, 12 şi 13 / Dennis Howitt, Duncan Cramer ; trad. de Andrei Popescu şi Cristina Popa. – Iaşi : Polirom, 2006

ISBN (10) 973-46-0294-2 ; ISBN (13) 978-973-46-0294-0

I. Cramer, Duncan
II. Popescu, Andrei (trad.)
III. Popa, Cristina (trad.)

004.42 SPSS

Printed in ROMANIA

Dennis Howitt
Duncan Cramer

Introducere în SPSS pentru psihologie

Versiunile SPSS 10, 11, 12 şi 13

Traducere de Andrei Popescu şi Cristina Popa

POLIROM
2006

COLLEGIUM

Psihologie



Cuprins

Introducere

Aplicaţia SPSS pentru ştiinţele sociale (*Statistical Package for the Social Sciences*, „pachet de programe statistice aplicabile ştiinţelor sociale") a fost iniţial elaborată în 1965 de către Universitatea Stanford din California. A fost primul utilitar pentru analiza datelor în acest domeniu, disponibil în universităţi şi în alte locaţii din întreaga lume. Progresele moderne în domeniul calculatoarelor au permis ca acest program să poată fi folosit şi pe calculatoarele personale. Datorită popularităţii sale, folosirea aplicaţiei SPSS este abilitatea cea mai uşor transferabilă dintre toate deprinderile necesare în cercetare. Utilizarea aplicaţiei SPSS, odată stăpânită, poate fi folosită practic în orice domeniu. În plus, aplicaţia SPSS este actualizată în mod constant, astfel încât cele mai recente metode statistice sunt incluse în fiecare versiune nouă.

Această carte nu necesită materiale adiţionale, este un mod de abordare pas cu pas a analizei statistice folosind SPSS pentru Windows şi este aplicabilă pentru versiunile 10,11,12 şi 13. Este destinată studenţilor şi cercetătorilor care doresc să analizeze date din psihologie, sociologie, criminologie, sănătate sau date similare.

Cartea îmbunătăţeşte modul de abordare din *A Guide to Computing Statistics with SPSS 11 for Windows* pentru a asigura un acces rapid către analiza computerizată a datelor folosind aplicaţia SPSS. Fiecare tehnică statistică este atent descrisă pas cu pas folosind capturi de ecran ale analizei statistice şi ale output-ului din aplicaţia SPSS. Atenţia utilizatorului este acum concentrată direct asupra capturilor de ecran, asupra semnificaţiilor şi importanţei fiecăreia. Cu alte cuvinte, abordarea folosită în acest manual se apropie mult de instruirea prin interacţiune directă, în limitele permise de un material scris. Persoanele cu puţină experienţă în utilizarea calculatoarelor sau cele fără experienţă vor putea să analizeze în scurt timp date complexe şi să evalueze semnificaţiile output-ului.

Majoritatea capitolelor au o structură asemănătoare. Paşii urmaţi de calculator (tastele care trebuie apăsate) sunt prezentaţi în ordinea exactă. Însă acesta nu reprezintă sfârşitul analizei datelor, aşadar există şi explicaţii referitoare la modul de interpretare şi prezentare a rezultatelor din SPSS. Structura comună este:

- Trăsăturile importante ale analizelor statistice din respectivul capitol sunt descrise într-o prezentare generală. Aceasta va fi suficientă, în majoritatea cazurilor, pentru ca utilizatorul să-şi facă o idee clară asupra locului şi momentului în care trebuie folosite acele tehnici.
- Urmează, apoi, o scurtă introducere referitoare la capitol şi la datele care vor fi analizate.
- Se dau exemple simple de tipuri de date, corespunzătoare fiecărei tehnici statistice. Aceste exemple îi permit utilizatorului să parcurgă pas cu pas calculele noastre şi să câştige încredere înainte de a trece la propriile date.
- Se prezintă, mai departe, pe cale vizuală, introducerea datelor pentru o anumită analiză statistică, iar explicaţiile sunt expuse în textul alăturat.

- Urmează o descriere pas cu pas, imagine cu imagine, a modului în care se realizează o anumită analiză statistică folosind aplicaţia SPSS pentru Windows.
- Output-ul statistic al SPSS-ului este ilustrat exact aşa cum apare pe monitor şi în textul (destinat listării) al analizei. Acest lucru este de importanţă crucială – output-ul SPSS-ului poate părea confuz şi neclar la început.
- Caracteristicile-cheie ale output-ului statistic sunt evidenţiate chiar în capturile de ecran, fiind însoţite de explicaţii simple asupra semnificaţiilor părţilor importante ale rezultatelor – aplicaţia SPSS fiind bine-cunoscută pentru complexitatea output-ului său.
- Se dau sugestii pentru prezentarea rezultatelor statistice în rapoarte, teze şi publicaţii. Printre acestea se regăsesc şi exemple ale modului de descriere a constatărilor cercetării şi de prezentare clară a tabelelor.

Această carte este bazată pe ultimele versiuni ale aplicaţiei SPSS pentru Windows (şi anume, versiunile 12 şi 13) ; este însă potrivită şi pentru versiunile 9, 10 şi 11, datorită asemănărilor existente. După această introducere urmează câteva observaţii care descriu diferenţele majore dintre versiuni. Deşi aplicaţia SPSS este actualizată o dată la câţiva ani, de obicei adaptarea la versiunea nouă a cunoştinţelor dobândite cu ajutorul unei versiuni anterioare se face cu uşurinţă.

Introducere în SPSS pentru psihologie (*Introduction to SPSS in Psychology*) reprezintă un unic îndrumar excelent pentru analiza datelor. Este o carte completă, de sine stătătoare, conţinând multe caracteristici care nu sunt prezentate în altă parte. Spre deosebire de alte cărţi despre SPSS, satisface nevoile studenţilor şi ale cercetătorilor la toate nivelele. Totuşi, este şi o parte componentă a unui pachet de cărţi metodologice scrise de aceiaşi autori, pachet conceput în ideea de a constitui o sursă exhaustivă şi de încredere. Cele trei volume din această serie sunt strâns legate între ele. Celelalte două sunt :

- *Introduction to Statistics in Psychology* (ediţia a III-a, Pearson Education : Harlow, 2005) – o iniţiere completă în statistică destinată tuturor studenţilor. Este alcătuită dintr-o introducere a noţiunilor-cheie ale statisticii în psihologie şi, deşi cartea este uşor de utilizat de către studenţi, prezintă în detaliu multe tehnici de nivel mediu şi avansat. Conţine capitole care tratează subiecte cum ar fi metaanaliza, care este rareori prezentată în alte texte statistice. Important este că structura acestui manual statistic este în strânsă legătură cu cartea de faţă. Deci oricine va urmări un capitol în manualul amintit va găsi, la locul potrivit, unul echivalent în această carte, cu detalii referitoare la analiza folosind SPSS. În acelaşi mod, orice persoană care utilizează această carte va putea găsi o prezentare în detaliu a tehnicii respective în manualul de statistică.
- *Introduction to Research Methods in Psychology* (Pearson Education : Harlow, 2005) – un manual de referinţă despre metodele de cercetare în psihologie. Acoperă atât metodele cantitative, cât şi pe cele calitative. Există capitole importante referitoare la scrierea rapoartelor, etica în psihologie şi parcurgerea literaturii de specialitate. Sunt acoperite toate aspectele elaborării studiilor experimentale, ale cercetărilor de teren, ale sondajelor şi chestionarelor şi se oferă îndrumări pentru colectarea de date şi realizarea unei analize de calitate. Există numeroase referiri reciproce între această carte şi cea amintită anterior.

Cu alte cuvinte, aceste trei cărţi oferă o introducere completă pentru realizarea cercetării în psihologie. Pot fi folosite independent sau în orice combinaţie.

Introducere în SPSS pentru psihologie poate fi utilizat împreună cu orice manual de statistică pentru susţinerea unei game variate de cursuri practice şi statistice. Aria tehnicilor statistice prezentate este largă, cuprinzând de la cele mai simple până la cele mai avansate şi importante tehnici statistice. Varietatea tehnicilor descrise şi relativa uşurinţă de folosire a aplicaţiei SPSS pentru Windows fac ca acest ghid să poată fi folosit ca metodă de predare a statisticii la nivel introductiv, intermediar şi avansat. Cartea este structurată în aşa fel încât procedurile statistice sunt descrise mai mult sau mai puţin în ordinea dificultăţii lor conceptuale. În general, modul de lucru cu SPSS este la fel de uşor pentru tehnicile statistice avansate, ca şi pentru cele simple.

Este foarte important să citiţi capitolul 1, pentru că explică introducerea datelor şi realizarea operaţiunilor de bază pe computer. Capitolele rămase pot fi uşor folosite ca materiale de sine stătătoare. Utilizatorii care nu dispun de suficient timp pentru a parcurge în amănunt acest ghid vor putea găsi suficiente detalii în capitolele-cheie pentru a putea realiza cu succes o analiză în SPSS. Tabelul 1.1 de la sfârşitul capitolului 1 explică rolul fiecărui capitol în parte, permiţând astfel cititorului să ajungă direct la acea parte.

Cei care vor parcurge cartea cu paşi mici vor profita din plin. Ei vor dispune de o privire de ansamblu mult mai bună asupra procedurilor de calcul din SPSS. Pentru majoritatea cititorilor, acest lucru este posibil în câteva ore, mai ales dacă au cunoştinţe de analiză statistică anterioare.

SPSS are un catalog extins de proceduri statistice – mult mai multe decât ar putea fi incluse într-un manual. Varietatea posibilităţilor poate să-l deruteze cu uşurinţă pe cititor, de aceea noi am selectat procedurile potrivite pentru majoritatea scopurilor. Uşurinţa şi rapiditatea aplicaţiei SPSS permit utilizatorilor mai avansaţi să exploreze posibilităţile existente, folosind meniurile şi ferestrele de dialog. Majoritatea utilizatorilor vor găsi prezentarea noastră mai mult decât suficientă.

Informaţiile şi analizele statistice efectuate în această carte corespund aproape întotdeauna cu cele din textul statistic înrudit, *Introduction to Statistics in Psychology* (ediţia a III-a, Pearson Education : Harlow, 2005), al aceloraşi autori. Când vom face referiri la această carte, o vom denumi *ISP*, menţionând capitolul sau numărul tabelului corespunzător.

Dorim să mulţumim tehnoredactorilor pentru realizarea capturilor de ecran, a casetelor de text şi a săgeţilor care le însoţesc, acestea fiind mult mai bune decât am fi putut noi să le facem.

Dennis Howitt
Duncan Cramer

Principalele diferențe dintre SPSS 13 și versiunile anterioare

SPSS 12

În ceea ce privește subiectul acestei cărți, diferențele esențiale dintre SPSS 12 și SPSS 13 sunt funcțiile „Compute Variable...", „Scatter/Dot..." și „Chart Editor". De asemenea, zonele din ferestrele aplicației SPSS 13 sunt acum umbrite.

În SPSS 12, fereastra de dialog „Compute Variable" are un singur ecran „Functions" din care pot fi selectate anumite opțiuni. „Scatter/Dot..." se numește acum „Scatter", fereastra de dialog „Scatter/Dot" se numește „Scatterplot" și opțiunea „Dot" nu există.

În SPSS 12, pentru a eticheta regiunile unei diagrame circulare și pentru a adăuga procentajul cazurilor în fiecare dintre ele, dați dublu clic oriunde în „Chart Editor", apoi dublu clic pe diagrama circulară (pentru a deschide fereastra de dialog „Properties"), selectați „Data Value Labels" (în fereastra de dialog „Properties"), selectați „Count" în ecranul „Contents", selectați „X"-ul roșu (pentru a introduce „Count" în ecranul „Available"), selectați numele variabilei (de exemplu, „Occupation"), selectați săgeata orientată în sus (pentru a introduce „Occupation" în ecranul „Contents"), selectați „Percent" și apoi săgeata orientată în sus (pentru a introduce „Percent" în ecranul „Contents"), selectați „Apply" și apoi „Close".

Pentru a introduce o linie de regresie într-un scatterplot, dați clic în tabelul din „Chart Editor", astfel încât cercurile din desen să fie evidențiate, selectați „Chart", apoi „Add Chart Element", selectați „Fit Line at Total" (care deschide fereastra de dialog „Properties").

Presupunând că opțiunea „Fit Line" este activă, selectați „Linear" (care, de obicei, este preselectat) și apoi apăsați „Close".

SPSS 11

Diferențele esențiale dintre SPSS 12 și SPSS 11 sunt valabile ca și în cazul versiunii SPSS 10. Acestea sunt relativ puține. În cazul versiunilor SPSS 11 și SPSS 10, denumirile variabilelor nu pot începe cu majusculă și sunt restricționate la 8 caractere.

Opțiunile „Data" și „Transform" nu sunt disponibile în ferestrele „Viewer" și „Output". Unele rezultate, cum ar fi corelația parțială și fiabilitatea, nu sunt organizate în tabele. „Chart Editor" funcționează diferit. Pentru a introduce o linie de regresie într-un scatterplot, dați dublu clic oriunde în scatterplot pentru a deschide

opțiunea „Chart Editor", selectați „Chart", selectați „Options..." (care deschide fereastra de dialog „Scatterplot Options"), selectați „Total" din meniul „Fit Line", iar apoi apăsați „OK".

SPSS 10

Între SPSS 10 și SPSS 11 sunt diferențe minore și abia observabile.

1 Noțiuni de bază despre introducerea datelor și analiza statistică folosind SPSS

Rezumat

- Acest capitol prezintă noțiuni de bază pentru operarea aplicației SPSS pe un computer personal. Descrie atât introducerea datelor, cât și salvarea fișierelor. Vor fi puține variații în accesarea SPSS-ului din locații diferite. Noțiunile de bază sunt destul de evidente și ușor de învățat.
- Sunt incluse câteva concepte statistice esențiale pentru a-i ajuta pe cercetători să planifice corespunzător analizele statistice.

1.1. Introducere

Versiunile SPSS 12 și SPSS 13 sau cele anterioare, cum ar fi SPSS 10 și SPSS 11, sunt de obicei disponibile pe calculatoarele din universități și colegii. Cei mai mulți utilizatori vor folosi SPSS-ul din aceste locații. În lumea întreagă, este de departe cea mai răspândită aplicație pentru analiza statistică. Prin urmare, învățarea utilizării SPSS-ului este o abilitate transferabilă, adeseori valoroasă pe piața muncii. Programul este folosit la toate nivelele, începând cu studenții și continuând cu cercetătorii specializați atât în numeroase domenii academice, cât și în situații practice. Un mare avantaj este acela că, odată însușite noțiunile de bază, SPSS-ul poate fi folosit la fel de ușor atât pentru analizele simple, cât și pentru cele complexe. Scopul acestui ghid este acela de a le oferi începătorilor o cale rapidă către facilitățile aplicației SPSS.

În zilele noastre majoritatea oamenilor sunt familiarizați cu operarea calculatoarelor personale (PC-urile). Începătorul absolut nu va fi dezavantajat, deoarece SPSS-ul este ușor de învățat. Utilizatorilor care sunt familiarizați, să spunem, cu procesarea de text li se va părea utilă aceasta în folosirea SPSS-ului – de exemplu, deschiderea programelor, deschiderea fișierelor și salvarea fișierelor. Experimentați fără teamă.

Din moment ce SPSS-ul este folosit adesea în cadrul universităților și colegiilor, multor utilizatori le va fi necesar un nume de utilizator și o parolă care se obțin de la instituția respectivă. De multe ori, documentația este pusă la dispoziție în momentul înregistrării la colegiu sau universitate. Dacă nu, aceasta este ușor de obținut.

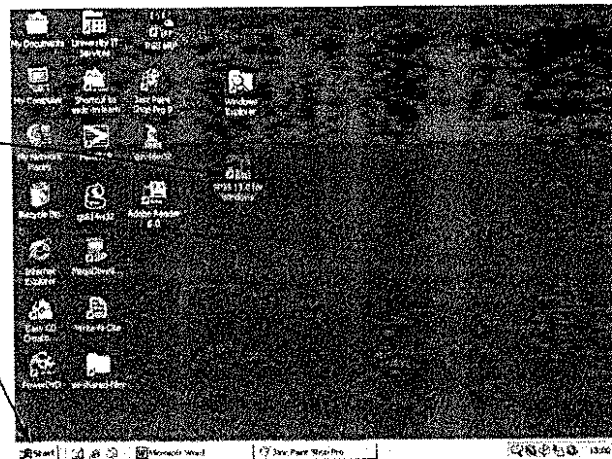
Pictograma SPSS-ului ar putea fi găsită printre cele de pe desktop sau în lista de programe ale Windows-ului. Fiecare instituție are propria manieră în ceea ce privește locația unde poate fi găsit SPSS-ul. Această carte are la bază versiunile 12 și 13 ale SPSS. Deși ele sunt ușor diferite față de versiunile precedente 10 și 11, instrucțiunile se aplică aproape în totalitate și pentru acestea din urmă.

1.2. Accesarea SPSS-ului

SPSS-ul destinat Windows-ului este accesat în general prin folosirea butoanelor şi meniurilor în conjuncţie cu clicuri ale mouse-ului. Aşadar, cea mai rapidă modalitate de învăţare este de a urma paşii prezentaţi şi capturile de ecran. Următoarea secvenţă de capturi de ecran este adnotată cu instrucţiuni şi marcată cu pasul 1, pasul 2 etc.

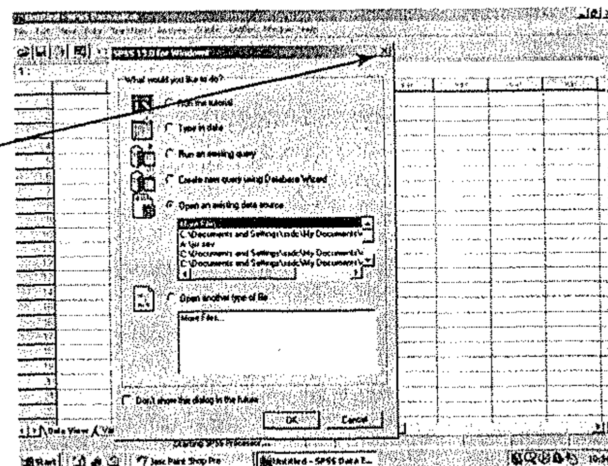
Pasul 1

Daţi dublu clic pe pictograma SPSS dacă aceasta apare pe ecran – dacă nu, apăsaţi butonul „Start" pentru a găsi lista de programe, deschideţi-o şi daţi clic pe SPSS.



Pasul 2

Această fereastră va apărea după câteva momente. Aţi putea alege oricare dintre opţiunile din fereastră. Totuşi, cel mai bine este să închideţi meniul frontal prin apăsarea butonului de închidere. E posibil ca acest meniu frontal să nu apară, deoarece poate fi închis definitiv.



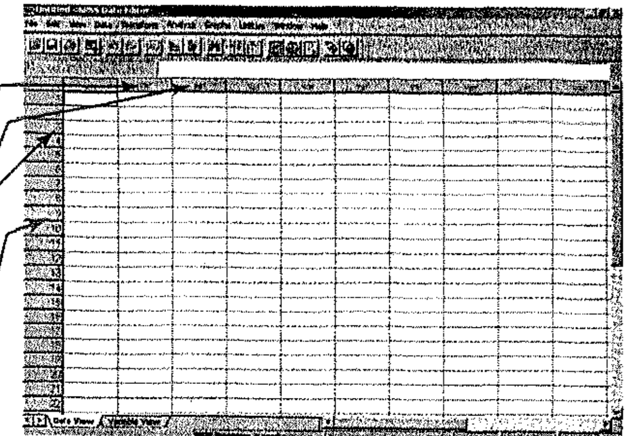
1.3. Introducerea datelor

Pasul 3

Data Editor al SPSS-ului poate fi acum vizualizat fără să fie acoperit de alte ferestre. Data Editor este un spaţiu de lucru cu linii şi coloane în care pot fi introduse date.

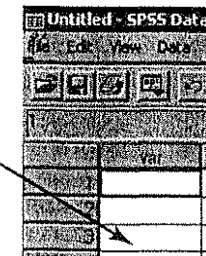
Coloanele sunt folosite pentru a desemna diferite variabile.

Liniile sunt cazurile sau indivizii despre care deţineţi date.



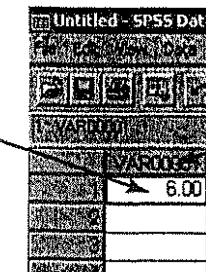
Pasul 4

Pentru a introduce date în SPSS selectaţi una dintre celule dând un clic pe celula respectivă – întotdeauna în SPSS există o singură celulă selectată.



Pasul 5

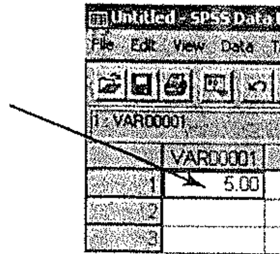
Introduceţi un număr folosind tastatura calculatorului. La apăsarea tastei „Enter" sau la selectarea cu mouse-ul a unei alte celule numărul va fi introdus în foaia de lucru la fel cum este arătat aici. Valoarea 6,00 este înregistrarea pentru primul rând (primul caz) al variabilei VAR00001.



Observaţi că această variabilă a primit automat un nume standard. Poate fi schimbat – daţi clic pe numele variabilei şi efectuaţi schimbarea.

Pasul 6

Corectarea erorilor – folosind mouse-ul, selectați celula unde este eroarea și dactilografiați valoarea corectată. La apăsarea tastei „Enter" sau deplasarea pe o altă celulă valoarea corectată va fi introdusă.



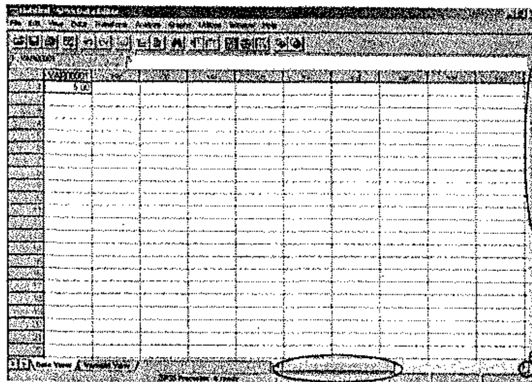
1.4. Deplasarea în cadrul unei ferestre folosind mouse-ul

Pasul 7

Puteți muta un rând sau o coloană la un moment dat apăsând pe butoanele cu săgeți ce se află la capătul barelor de derulare situate vertical și orizontal.

Pentru deplasări ample, mișcați barele de derulare verticale și orizontale pentru deplasarea în cadrul paginii.

Poziția relativă a barelor de derulare indică poziția relativă în cadrul fișierului.

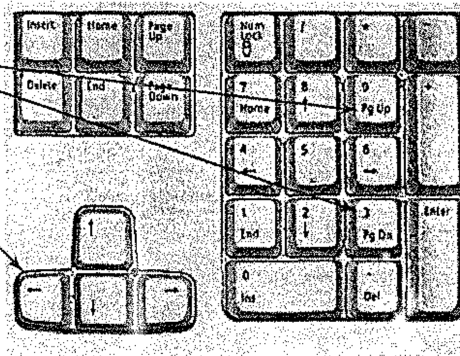


1.5. Deplasarea în cadrul unei ferestre folosind tastatura și mouse-ul

Pasul 8

Vă puteți deplasa cu o pagină mai sus sau mai jos pe ecran apăsând tastele „Pg Up" și „Pg Dn".

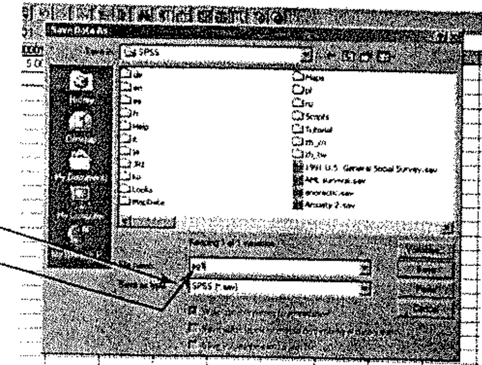
Săgețile de la tastatură mută cursorul un spațiu sau un caracter, în funcție de direcția săgeții.



1.6. Salvarea datelor

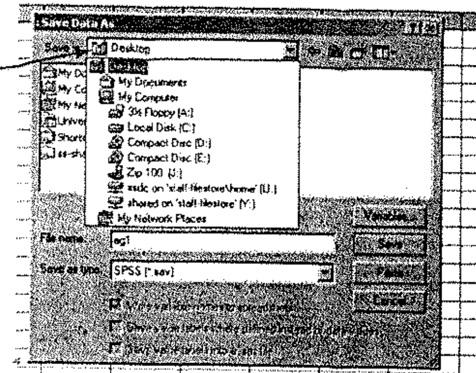
Pasul 9

Selectând „File" și apoi „Save As..." veți putea salva datele sub forma unui fișier. Fișierul cu datele salvate primește automat de la SPSS extensia „.sav". Este utilă folosirea unor nume de fișiere distincte, cum ar fi „eg1", pentru a face conținutul lor cât mai clar.



Pasul 10

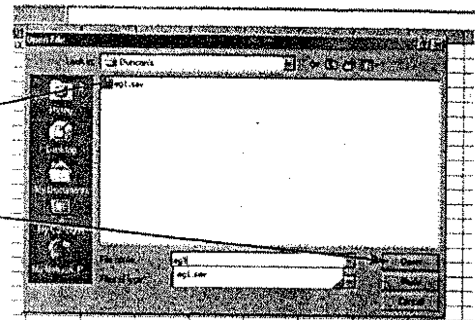
Pentru a alege locația unde fișierul de date va fi salvat, indicați calea în meniul „Save in:". Folosiți săgeata pentru a ajunge la locația dorită.



1.7. Deschiderea unui fișier de date

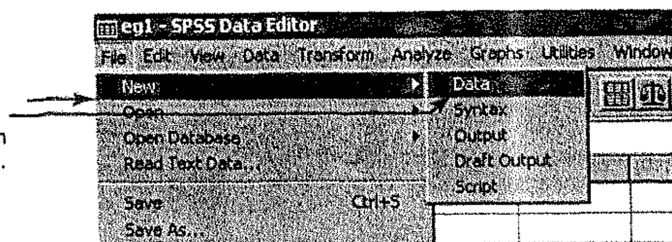
Pasul 11

Pentru a deschide un fișier existent apăsați pe „File", „Open", „Data", „Look in:", dacă fișierul nu este în meniul „Open File" (ca în cazul în care dacă tocmai l-ați salvat), scrieți numele fișierului („eg1") și apoi apăsați „Open".



Pasul 12

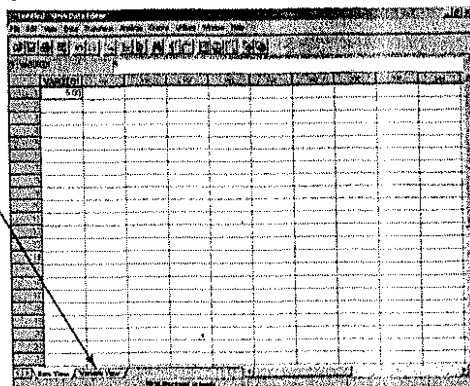
Pentru a deschide un fişier nou apăsaţi pe „File", „New" şi apoi „Data". Acest fişier poate fi salvat aşa cum a fost arătat la pasul 9.



1.8. Folosirea „Variable View" pentru crearea şi etichetarea variabilelor

Pasul 13

Apăsarea meniului „Variable View", situat în josul paginii, schimbă fereastra „Data View" (spaţiul de lucru pentru introducerea datelor) în altă fereastră în care pot fi introduse, convenabil, informaţii despre variabilele pe care le-aţi creat.

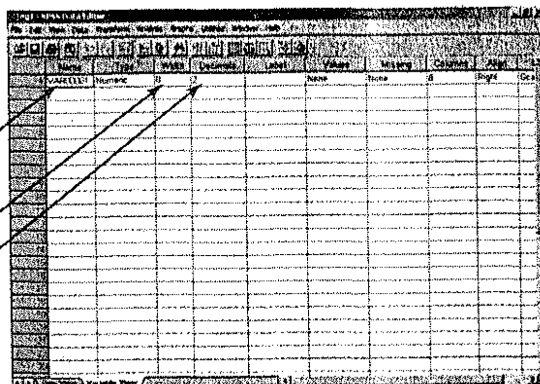


Pasul 14

Acesta este spaţiul de lucru „Variable View". În acest caz, o variabilă este deja înregistrată. A fost introdusă la pasul 5. Poate fi redenumită şi pot fi adăugate alte variabile destul de uşor doar selectând celula corespunzătoare şi tastând numele variabilei.

Aici modificaţi lăţimea coloanei de date.

Aici modificaţi numărul de zecimale.



Pasul 15

În versiunile 12 şi 13, faţă de cele anterioare, nu există limită pentru lungimea numelui unei variabile. Versiunile anterioare erau limitate la 8 caractere. Selectaţi o celulă din coloana „Name" şi dactilografiaţi un nume de *variabilă* diferit. Restul coloanelor vor primi valori implicite care pot fi schimbate ulterior. Aceste variabile redenumite şi nou definite vor apărea în „Data View" când această fereastră va fi selectată.

Acesta este numărul de zecimale care va apărea pe ecran – în calcule este folosită valoarea zecimală fără aproximări.

Pasul 16

Este important de observat că şi alte coloane pot fi modificate uşor.

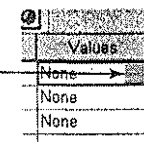
Coloana „Label" permite atribuirea variabilelor unor nume mai lungi decât este posibil în coloana „Variable name". Pur şi simplu scrieţi un nume mai lung în celule. Acest lucru nu mai este atât de important din moment ce versiunea SPSS 12 permite utilizarea unor nume de variabile mult mai lungi.

Capitolul 15 se ocupă de valorile lipsă (coloana „Missing").

Coloana „Values" permite definirea unor categorii sau valori ale unor variabile nominale (sau categoriale), cum ar fi genul. Consultaţi pasul 17 pentru detalii. Este recomandat să existe valori în coloana „Values" pentru toate variabilele nominale.

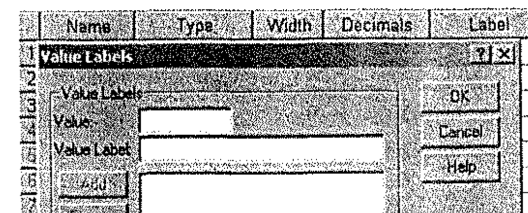
Pasul 17

Apare acest buton. Daţi clic pe el.



Pasul 18

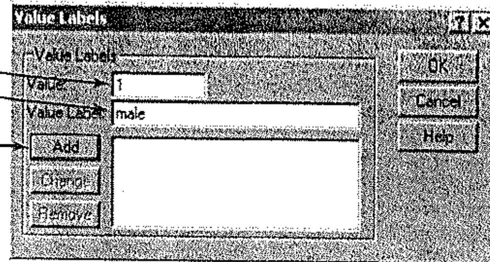
Va apărea această fereastră. Urmăriţi în continuare paşii. Vă vor arăta cum „male" şi „female" sunt introduse folosind codul „1" pentru bărbat şi „2" pentru femeie.



Pasul 19

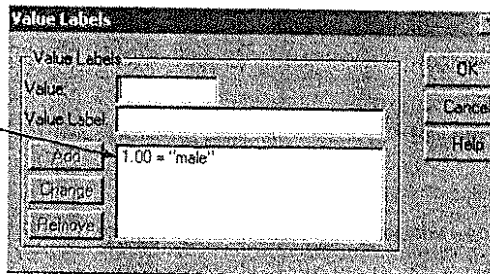
Scrieți „1" în dreptul câmpului „Value" și „male" în dreptul câmpului „Value Label:".

Apoi apăsați „Add".



Pasul 20

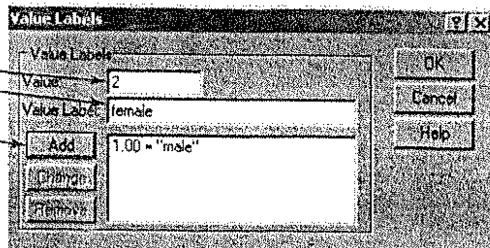
Această operație transferă informația în câmpul aflat dedesubt.



Pasul 21

Acum scrieți „2" în dreptul câmpului „Value" și „female" în dreptul câmpului „Value Label:".

Apoi apăsați „Add".

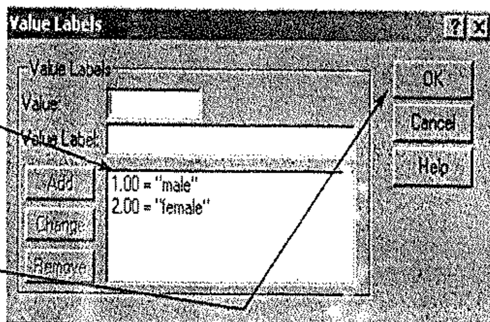


Pasul 22

Această operație transferă informația în căsuța mai mare.

Este bine să etichetați valorile în acest fel.

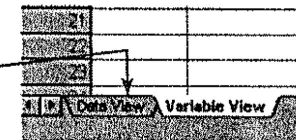
Apăsați „OK" pentru a închide fereastra.



1.9. Mai multe detalii despre „Data View"

Pasul 23

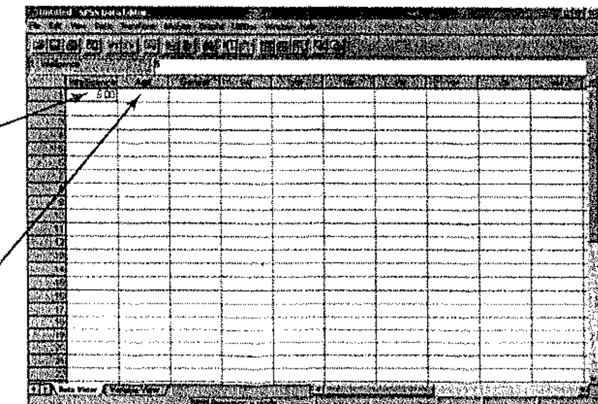
Pentru întoarcere la „Data View" apăsați acest tab din colțul din stânga jos al ferestrei.



Pasul 24

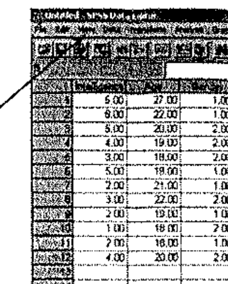
Așa arată „Data View" acum.
Pot fi introduse informațiile pentru toate variabilele și cazurile. Nu uitați că valoarea 5,00 a fost introdusă anterior împreună cu numele variabilei.
Putem începe introducerea informațiilor în întregime.

Pentru a introduce o informație, selectați o celulă, introduceți numărul, și apoi apăsați tasta „Enter". Veți ajunge în celula aflată sub cea selectată anterior, care va deveni astfel noua celulă selectată.



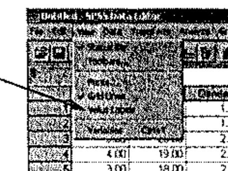
Pasul 25

Această captură prezintă felul în care arată o foaie tipică de calcul cu informații. Observați cum valorile pentru gen („gender") sunt codate cu 1,00 și 2,00. Este posibil ca în locul acestor valori să afișăm valorile etichetelor. Dați clic pe butonul „View" din bara de sarcini.



Pasul 26

Apoi dați clic pe „Value Labels".



Pasul 27

Acum valorile sunt date ca „male", „female" aşa cum au fost codate de la paşii 19 până la 21.

Pasul 28

Sunt multe opţiuni disponibile pentru dumneavoastră, inclusiv analize statistice. Câteva dintre aceste opţiuni sunt prezentate aici.

Selectaţi „Data" pentru a introduce noi variabile, noi cazuri, pentru a selecta cazuri şi pentru alte manipulări de date.

Selectaţi „Window" pentru a comuta între foile de calcul cu date şi pe cele cu rezultate.

Selectaţi „Transform" pentru a modifica datele – cum ar fi modificarea valorilor şi calculul combinat de variabile.

Selectaţi „Analyze" pentru a avea acces la toată gama de calcule statistice pe care le realizează SPSS.

Selectaţi „Graphs" pentru diagrame cu bare, scatterplot-uri şi multe alte metode de reprezentare grafică.

1.10. Un calcul statistic simplu

Pasul 29

Pentru a calcula media de vârstă urmaţi etapele următoare:

Daţi clic pe „Analyze".

Selectaţi „Descriptive Statistics".

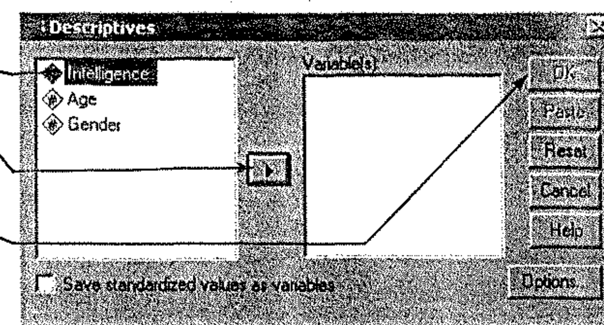
Selectaţi „Descriptive...".

Pasul 30

Apare această fereastră. Selectaţi „Intelligence".

Daţi clic pe butonul ▶ pentru a muta „Age" în căsuţa „Variable(s):".

Daţi clic pe „OK".



1.11. Output-ul SPSS-ului

Pasul 31

Fereastra „Data Editor" este înlocuită de output-ul SPSS-ului. Acest tabel apare pentru analiza realizată anterior.

Media scorului pentru inteligenţă este încercuită.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Intelligence	12	1.00	8.00	3.6667	1.92275
Valid N (listwise)	12				

Vestea bună este că orice persoană care poate urma paşii de mai sus nu ar trebui să aibă nici o dificultate în efectuarea marii majorităţi a analizelor statistice disponibile în aplicaţia SPSS, cu ajutorul acestei cărţi. Merită să petreceţi în jur de o oră exersând cu SPSS. Veţi vedea că aceasta este cea mai rapidă modalitate de a învăţa.

1.12. Concepte statistice esenţiale în analizele cu ajutorul SPSS-ului

Elementele de statistică sunt destul de simple. Problema apare la asamblarea elementelor. Nimeni nu poate deveni peste noapte expert în analiză statistică, dar cu un bagaj mic de cunoştinţe, un cercetător neexperimentat poate întreprinde analize sofisticate. Abilităţile matematice au un rol mic în analiza statistică a datelor.

Sunt câteva concepte de bază pe care cercetătorii trebuie să le înţeleagă înainte de a trece la analizele SPSS. Acestea includ:

- *Variabila.* O variabilă este orice concept care poate fi măsurat şi care variază. Variabilele sunt în general invenţii ale cercetătorilor şi variază enorm de la un studiu la altul. Sunt câteva variabile standard, cum ar fi vârsta sau genul, care sunt măsurate frecvent. De obicei, variabilele sunt specifice unui segment particular al studiului. Variabilele apar în analizele SPSS pe coloane în foile de calcul cu date.

- *Cazurile*. Un caz este doar un membru al unei mostre. În psihologie un caz este de obicei o persoană (un individ care participă la studiu). Cazurile (în mod normal) apar în analizele SPSS ca fiind rândurile unei foi de calcul cu date.
- *Tipurile de variabile*. Pentru toate scopurile practice, variabilele pot fi clasificate ca fiind de două tipuri :
  - *Scoruri*. Unele variabile sunt scoruri. Vorbim despre scoruri atunci când o valoare numerică este atribuită unei variabile, pentru fiecare caz în parte din eșantion. Această valoare numerică indică cantitatea sau valoarea caracteristicii (variabilei) în cauză. Astfel, vârsta este o variabilă numerică deoarece valoarea-scor indică o cantitate în creștere a variabilei-vârstă. Unii ar putea cataloga acest tip de variabilă drept cantitativă.
  - *Variabile nominale sau categoriale*. Unele variabile sunt măsurate clasificând cazurile în unul sau mai multe categorii. Acestea sunt cunoscute drept variabile nominale ori categoriale. De exemplu, genul are două categorii denumite bărbat și femeie. Naționalitatea este un alt exemplu : engleză, galeză, irlandeză și scoțiană sunt naționalitățile popoarelor din Marea Britanie. Aceste variabile nu au nici o implicație numerică. A spune că o persoană este de naționalitate scoțiană este echivalent cu punerea ei în respectiva categorie nominală. Unii ar putea privi lucrurile calitativ. Există un risc de confuzie – categorii precum genul sunt introduse în SPSS folosind numere diferite pentru categorii diferite. De exemplu, variabila-gen are două categorii – bărbații ar putea fi reprezentați prin numărul „1" și femeile prin numărul „2" (sau viceversa). Numerele sunt folosite arbitrar. Este foarte important a nu se confunda aceste numere, care reprezintă numai codarea diferitelor categorii. Pentru acest motiv este important să etichetați explicit valorile diferite ale variabilelor nominale în foile de calcul SPSS. Acest lucru poate fi făcut cu ușurință, așa cum a fost arătat la paginile 7 și 8.

  Câteodată variabilele sunt clasificate ca fiind nominale, ordinale, de interval și de raport. Acest lucru este de interes conceptual, însă are o importanță practică redusă în selectarea statisticii corespunzătoare.
- Este important de hotărât pentru fiecare variabilă în parte dacă este variabilă nominală (categorială) sau variabilă cantitativă. Dacă sunteți începător, scrieți o listă cu variabilele și clasificați-le pe fiecare în parte. În cele din urmă veți face asta fără prea mare bătaie de cap. Tehnicile statistice care corespund variabilelor cantitative nu se potrivesc totdeauna pentru cele calitative (și viceversa). Așadar, este convenabil de calculat media (cantitativă) pentru orice variabilă cantitativă (de exemplu, media de vârstă). Pe de altă parte, nu este convenabil să se calculeze media pentru variabile care denumesc o categorie. Ar fi un nonsens să spunem că media naționalității este 1,7 din moment ce naționalitatea nu este un scor. Problema este că SPSS-ul lucrează cu numere în foile de prelucrare a datelor și nu poate ști dacă este vorba despre scoruri sau coduri pentru categorii diferite.
- Există două tipuri principale de tehnici statistice – statistici *descriptive* și *inferențiale* :
  - Statisticile descriptive descriu mai ales caracteristicile principale ale variabilelor individuale. Deci calculul vârstei medii pentru un eșantion de locuitori este un exemplu de statistică descriptivă. Numărarea locuitorilor de naționalitate engleză este un alt exemplu de statistică descriptivă. Statistici bivariate sunt folosite atunci când relația dintre două (sau mai multe) variabile este descrisă.

  - Statisticile inferențiale reprezintă un aspect cu totul diferit al statisticii. Se aplică numai pentru a răspunde la întrebarea dacă ne putem baza pe rezultatele obținute pe un eșantion de cazuri. Folosirea eșantioanelor este caracteristică pentru majoritatea studiilor moderne. Problema cu eșantioanele este că unele dintre ele nu sunt similare cu populația din care au fost extrase. Expresiile „statistic semnificativ" și „statistic nesemnificativ" indică dacă o tendință a datelor poate fi acceptată ca substanțială (statistic semnificativ) sau insuficient de substanțială pentru a ne baza pe ea (statistic nesemnificativ). Obținerea semnificației statistice este puțin probabil să fie rezultatul întâmplării. A se vedea îndrumarul statistic al autorilor, D. Howitt și D. Cramer, *Introduction to Statistics* (ediția a III-a, Pearson Educational : Harlow, 2005), pentru o prezentare în detaliu a înțelesului „semnificației statistice", deoarece este dificil de explicat cu acuratețe doar în câteva cuvinte.
  - Fiecărei statistici descriptive îi corespunde o statistică inferențială. De exemplu, coeficientul de corelație este un indicator statistic descriptiv indicând direcția și intensitatea relației dintre două variabile. Asociat acestui coeficient este statistica inferențială – semnificația coeficientului de corelație. Statistica descriptivă este importantă pentru înțelegerea tendinței datelor – statistica inferențială se ocupă de încrederea pe care o putem avea în rezultatele obținute.
- Cercetătorii trebuie, de asemenea, să țină cont de cele *două* tipuri diferite de modele de cercetare – acelea care folosesc *măsurători relaționate* sau *măsurători nerelaționate*. De asemenea, măsurătorile relaționate pot fi numite măsurători corelate sau măsurători pereche. Apoi, măsurătorile nerelaționate pot fi numite măsurători necorelate sau măsurători independente. Acești termeni sunt folosiți cu precădere atunci când valorile medii ale numerelor sunt comparate pentru două sau mai multe eșantioane de date :
  - Când mediile obținute pentru un singur eșantion de indivizi se compară cu două (sau mai multe) măsurători ale aceleiași variabile (de exemplu, cele preluate la momente diferite de timp), atunci acesta este un model de măsurători relaționate.
  - Când mediile pentru două eșantioane diferite de participanți sunt comparate cu o variabilă, acesta este un model nerelaționat.
  - Când două (sau mai multe) grupe de participanți au fost potrivite cu atenție astfel încât *seturi* de participanți în cele două (sau mai multe) condiții sunt similare în unele aspecte, atunci acesta este, de asemenea, un model relaționat. În acest caz, membrii fiecărui set sunt tratați ca fiind o aceeași persoană. În mod normal, un cercetător va ști dacă participanții au fost asociați în seturi, deoarece acest lucru necesită efort din partea cercetătorului. De exemplu, acestea din urmă trebuie să decidă care sunt caracteristicile de asociere a seturilor, iar apoi va alege indivizi pe baza similitudinilor caracteristicilor, și (adeseori) trebuie să distribuie participanți în eșantioane diferite (condiții) în special în cercetarea experimentală.

Scopul principal pentru folosirea modelelor relaționate este reducerea variabilității cauzate de eșantionare.

Aproape fără excepție, cercetătorul va folosi o varietate de tehnici pentru aceleași date. Din fericire, odată introduse datele, în multe cazuri, analiza acestora durează ceva mai mult de un minut.

1.13. Ce teste trebuie folosite

O nevoie frecventă este aceea de a şti cum se alege tehnica statistică corespunzătoare pentru date. De-a lungul timpului, autorii de manuale statistice au lucrat în vederea simplificării procesului de alegere. S-a lucrat, în mare măsură, la elaborarea de foi de lucru care indică tipul de statistică corespunzător pentru diferite tipuri de date. Dacă veţi dori o asemenea abordare, atunci aveţi la dispoziţie aceste site-uri care vă vor ghida de-a lungul procesului de alegere:

http://www.orthoteers.co.uk/Nrujp~33lm/Orthstatstests.htm
http://www.wtc.edu/online/bcanada/choose2.htm
http://www.members.aol.com/statware/pubpage.htm
http://www.socialresearchmethods.net/selstat/ssstart.htm
http://www.graphpad.com/www/Book/Choose.htm
http://www.whichtest.info/
http://www.google.com/search?q=choosing+statistical+tests&h1=en&lr=&ie=UTF-8&safe=off&start=10&sa=N
http://www.google.com/search?q=choosing+statistical+tests&h1=en&lr=&ie=UTF-8&safe=off&start=20&sa=N

Pentru statistica de bază este probabil o abordare folositoare. Dificultatea crescând, face ca modelele de cercetare, chiar şi pentru proiectele studenţilor, să fie foarte variate şi destul de complexe. Cândva, psihologia era aproape numai o ştiinţă bazată pe subiecte de laborator, care se concentra pe experimente aleatorii. Psihologii folosesc în continuare astfel de experimente, dar metodele s-au extins cu mult, în acest fel extinzându-se şi nevoia de cunoştinţe statistice. Astfel, există o limită pentru măsura în care o foaie de lucru simplă sau o schemă logică pot ajuta cercetătorii să selecteze analiza statistică corespunzătoare.

O greşeală fundamentală a cercetătorilor începători este aceea de a presupune că analiza datelor este condusă înainte de toate de statistică. Este mai mult decât corect să privim statistica drept un instrument care aduce fineţe scopului de bază al cercetării – pentru a răspunde la întrebarea cercetătorului cu privire la studiu. Numai cercetătorul poate şti în întregime ceea ce vrea să obţină prin studiul său – ce probleme doreşte să fie rezolvate prin intermediul colectării datelor şi al analizei lor. Dacă cercetătorul nu înţelege clar ce vrea să obţină prin intermediul studiului său, statistica nu-i poate fi de ajutor. Adeseori, când suntem consultaţi pentru sfaturi privind statistica, aflăm că întâi de toate trebuie să clarificăm obiectivele cercetării – abia apoi să încercăm să desluşim felul în care cercetătorii au gândit utilizarea datelor corecte. Acestea *nu* sunt chestiuni statistice, ci probleme referitoare la dezvoltarea ideilor studiului şi planificarea corespunzătoare a achiziţiei de date. Deci prima chestiune ţine de evidenţierea întrebărilor pentru care se caută un răspuns, cu datele existente. Prea des, atingerea scopului cercetării este pierdută în mulţimea chestiunilor practice. Următoarele rânduri vor ajuta la clarificarea rolului statisticii în cercetare:

- Multe dintre aspectele importante ale analizei datelor nu necesită altceva decât înţelegerea valorilor mediilor şi ale frecvenţelor. Acestea sunt des întâlnite printre rezultatele din SPSS. Multe dintre întrebările studiului pot primi răspuns pe baza analizei diferenţelor de medii între eşantioane, tabele de asociere sau diagrame scatter. Este folositor să ne punem singuri întrebarea cum ar putea fi găsite răspunsuri la întrebările studiului doar utilizând asemenea abordări de bază. Prea des complexitatea rezultatelor statistice obţinute devine centrul atenţiei, ceea ce conduce la confuzie în ceea ce priveşte legătura dintre date şi scopurile cercetării. Nu este uşor să ne concentrăm asupra problematicii studiate, fără să fim, în acelaşi timp, atraşi în direcţii greşite.
- Analizele statistice sunt construite dinamic de către cercetător. De obicei nu există o singură analiză statistică corectă pentru un set de date, ci o gamă de alternative în mod egal acceptate. Cercetătorul va trebui să ia multe decizii de-a lungul procesului de analiză a datelor – unele dintre ele vor trebui justificate cu atenţie, iar altele pot fi aproape arbitrare. Cercetătorul se ocupă de analiza datelor – statistica este instrumentul cercetătorului. Analiza nu trebuie să fie un tur de forţă al statisticii, ci să se lase condusă de întrebările care au necesitat colectarea datelor de la bun început. Nu există nici o scuză – dacă aţi colectat datele, trebuie să ştiţi de ce aţi făcut acest lucru.
- Cu cât citiţi în întregime mai multe studii, cu atât mai bine veţi înţelege cum poate fi folosită statistica într-un domeniu particular de cercetare. Există foarte puţine studii care să nu aibă nici o legătură cu alte studii. Care sunt metodele statistice folosite de cercetători pentru domeniul ales? Dacă ştiţi ce tehnici sunt folosite în general, acest lucru vă poate fi de mare folos atunci când veţi decide ce anume trebuie luat în consideraţie.

Tabelul 1.1 oferă o imagine asupra stilurilor de analiză pe care cercetătorii ar putea dori să le aplice în cazul propriilor date şi a secţiunilor în care sunt descrise în detaliu aceste tehnici statistice.

Tabelul 1.1. *Principalele tipuri de analiză şi procedurile SPSS sugerate*

Tipul/obiectivul analizei	Proceduri sugerate	Capitolul
Toate tipurile de studii	Statistici descriptive, tabele şi diagrame	2-6
Evaluarea relaţiei dintre două variabile	Coeficientul de corelaţie	7
	Regresia	8
Compararea a două seturi de scopuri pentru depistarea diferenţelor	Testul *t* pentru eşentioane independente	13
	Testul *ratio F*	19
	Testul *t* pentru eşantioane perechi	12
	ANOVA pentru eşentioane independente	20
	ANOVA pentru eşantioane perechi	21
	Testul Mann-Whitney	18
	Testul Wilcoxon al rangurilor pereche	18
Compararea mediilor a două sau mai multe seturi de scopuri	ANOVA pentru eşantioane independente	20
	ANOVA pentru eşantioane perechi	21
	Comparaţiile multiple	23

Tipul/obiectivul analizei	Proceduri sugerate	Capitolul
Experimente complexe etc., având *două* sau mai multe variabile independente şi *o* variabilă dependentă	ANOVA bifactorială sau multifactorială	22
- dacă efectuaţi măsurători relaţionate *şi* nerelaţionate	Modelul ANOVA mixt	24
- dacă există alte variabile care pot afecta valorile variabilei dependente	Analiza covarianţei	24
Eliminarea celei de-a treia variabile care poate afecta coeficientul de corelaţie	Corelaţia parţială	26
Identificarea predictorilor pentru o variabilă numerică	Regresia simplă	8
	Regresia multiplă stepwise	28
	Regresia multiplă ierarhică	29
	Analiza logliniară	31
Identificarea predictorilor pentru o variabilă categorială	Regresia logistică multinomială	32
	Regresia logistică binomială	33
Analizarea unui chestionar	Analiza factorială	27
	Coeficientul alpha de validitate	30
	Metoda split-half	30
	Recodificarea	16
	Procesarea variabilelor noi	17
Compararea datelor de frecvenţă	Testul chi-square	14
	Testul Fisher	14
	Testul McNemar	14
	Analiza logliniară	31
Codificarea datelor deschise cu ajutorul evaluatorilor	Coeficientul Kappa	30

2 Descrierea variabilelor
Tabele şi diagrame

Rezumat

- Tabelele şi diagramele au menirea de a evidenţia rapid şi eficient trăsăturile importante ale datelor. Complexitatea, ca scop în sine, nu este o trăsătură utilă a tabelelor şi diagramelor bune.
- Tabelele şi diagramele clare şi semnificative sunt deosebit de importante pentru analiza statistică şi întocmirea rapoartelor. Cu mici excepţii, toate analizele de date le utilizează pentru a permite interpretarea distribuţiilor de variabile. În acest capitol vom prezenta tehnicile de bază care permit construirea de tabele şi diagrame pentru a descrie distribuţia variabilelor prezentate una câte una.
- Toate tabelele şi diagramele trebuie să fie denumite şi etichetate clar. În funcţie de tabelul sau diagrama în cauză, axele orizontale şi verticale trebuie să fie etichetate, barele identificate, scala marcată ş.a.m.d. Acest lucru trebuie făcut cu mare grijă. Vorbind la modul general, tabelele şi diagramele din SPSS necesită o muncă considerabilă pentru a dobândi eficienţă.
- Tabelele de frecvenţă numără, ori de câte ori apar, diferite valori ale variabilelor. Un exemplu simplu ar fi numărarea bărbaţilor şi a femeilor care participă la studiu. Tabelele trebuie să fie relativ simple şi în mod obişnuit necesită fie ca variabila să aibă un număr mic de valori, fie ca un număr de valori diferite ale variabilei să fie grupate.
- Diagramele circulare sunt căi eficiente şi simple de a prezenta frecvenţele. Totuşi, ele sunt utile doar dacă variabila ilustrată deţine un număr mic de valori distincte. Diagramele circulare sunt rareori folosite în publicaţii deoarece consumă spaţiu, deşi sunt bune pentru prezentările din cadrul conferinţelor şi pentru prelegeri.
- O diagramă cu bare (bar chart) poate fi folosită în circumstanţe asemănătoare celor pentru diagrama circulară, dar poate face faţă unui număr mai mare de valori ale variabilelor fără să devină prea înghesuită. Frecvenţele diferitelor valori ale variabilelor sunt reprezentate cu bare separate fizic şi de înălţimi care variază în funcţie de frecvenţa acelei categorii.
- O histogramă este similară unei diagrame cu bare, dar se foloseşte mai degrabă pentru scoruri numerice decât pentru categorii. Astfel, barele dintr-o histogramă apar în ordinea mărimii variabilelor numerice pe care le reprezintă. Într-o histogramă, barele *nu* sunt separate de spaţii. Adeseori o histogramă va necesita ca domeniul variabilelor numerice afişate de fiecare bară în parte să fie modificat pentru a maximaliza utilitatea şi claritatea diagramei. Acest lucru poate fi făcut recodificând variabilele (capitolul 16), dar SPSS-ul permite ca acest lucru să fie făcut în „Chart Editor". Realizarea unor asemenea tabele este una dintre cele mai dificile sarcini în SPSS.

Tabelul 2.1. *Ocupația participanților la studiu exprimată în frecvență și în procentaj de frecvență*

Occupation	Frequency	Percentage frequency
Nuns	17	21.25
Nursery teachers	3	3.75
Television presenters	23	28.75
Students	20	25.00
Other	17	21.25

SPSS este folosit în general pentru a rezuma datele neprelucrate, dar poate utiliza și date care au fost deja rezumate, cum ar fi cele prezentate în tabelul 2.1 (*ISP*, tabelul 2.1).

Cu alte cuvinte, din moment ce tabelul 2.1 este bazat pe 80 de persoane, datele vor ocupa 80 de celule ale unei singure coloane în fereastra „Data Editor", și fiecare ocupație va fi codată cu un număr diferit, deci „Nuns" ar putea fi codat „1", „Nursery teachers" cu „2" ș.a.m.d. Astfel, sunt necesare 17 rânduri care conțin „1" pentru a reprezenta categoria „Nuns", 3 rânduri care conțin „2" pentru a reprezenta categoria „Nursery teachers" ș.a.m.d. Totuși, este posibil de dus la bun sfârșit anumite analize cu date rezumate având în vedere aprecierea convenabilă a categoriilor, după numărul sau frecvența cazurilor existente.

2.1. Realizarea tabelelor de frecvență prin ponderare

Pare o practică bună definirea variabilelor în fereastra „Variable View" din „Data Editor" înainte de a introduce datele în „Data View", pentru că putem înlătura câmpurile zecimale acolo unde nu sunt necesare. Deci vom face mai întâi acest lucru. Dacă preferați să introduceți datele mai întâi puteți proceda și în acest fel.

Pentru acest tabel avem nevoie de două coloane în „Data View". Una pentru a arăta genul categoriilor. Alta pentru a introduce frecvențele acestor categorii. În „Variable View" variabilele sunt prezentate ca rânduri. Din moment ce vom introduce date în coloane ne vom referi la aceste rânduri ca fiind coloane.

Pasul 1

Selectați „Variable View" din „Data Editor".

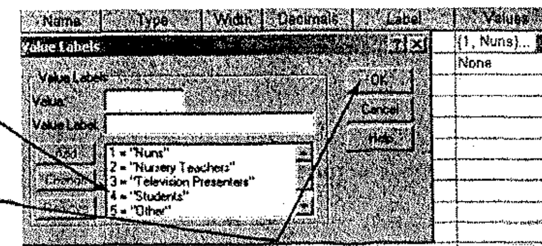
Denumiți primele două coloane: „Occupation" și „Freq". Înlăturați cele două câmpuri zecimale.

Selectați partea dreaptă a celulei pentru „Values" ale „Occupation".

Pasul 2

Etichetați cele cinci valori pentru „Occupation" așa cum este descris în pașii 18-20 din capitolul 1.

Apăsați „OK".



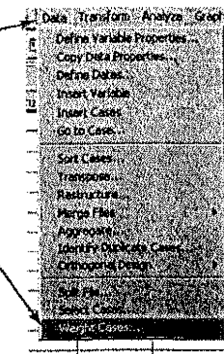
Pasul 3

Selectați „Data View". Introduceți datele așa cum este arătat.

Occupation	Freq
1	17
2	3
3	23
4	20
5	17

Pasul 4

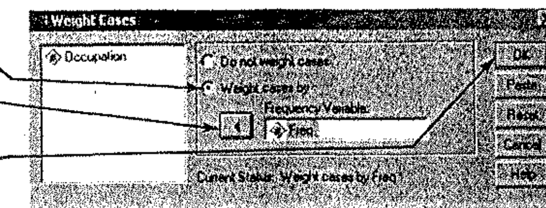
Selectați „Data" și „Weight Cases...".



Pasul 5

Selectați „Freq", „Weight cases by" și apoi apăsați butonul ◄ pentru transfer în căsuța variabilei „Frequency Variable:".

Apăsați „OK"



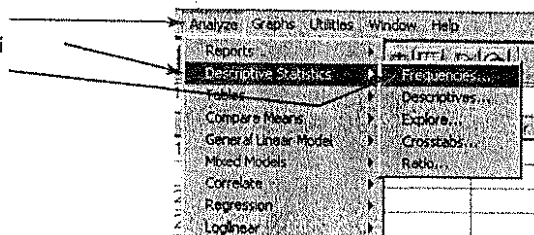
Cazurile sunt acum ponderate în funcție de frecvențe așa cum este arătat de mesajul „Weight On" din colțul dreapta jos al ferestrei „Data Editor".



2.2. Frecvența procentuală

Pasul 1

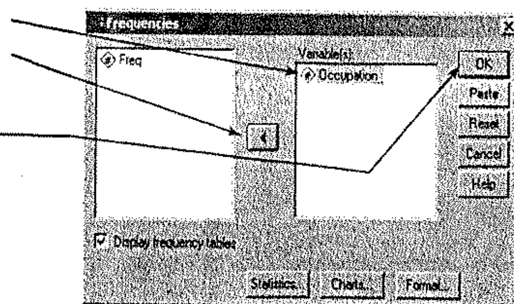
Selectați „Analyze", „Descriptive Statistics" și „Frequencies...".



Pasul 2

Selectați „Occupation" și apoi apăsați butonul ◀ pentru transfer în căsuța de variabile „Frequency Variable(s):".

Apăsați „OK".



2.3. Interpretarea output-ului

Prima coloană a acestui tabel conține valorile pentru cele cinci categorii. Pentru a le schimba cu etichete, ceea ce este mai semnificativ, parcurgeți următorii trei pași și reluați analiza așa cum este descris în tabelul alăturat.

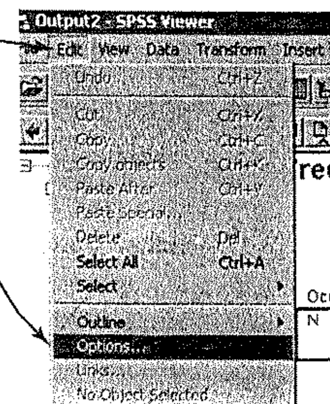
Ocupație

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	17	21.3	21.3	21.3
	2	3	3.8	3.8	25.0
	3	23	28.8	28.8	53.8
	4	20	25.0	25.0	78.8
	5	17	21.3	21.3	100.0
	Total	80	100.0	100.0	

2.4. Etichetarea valorilor

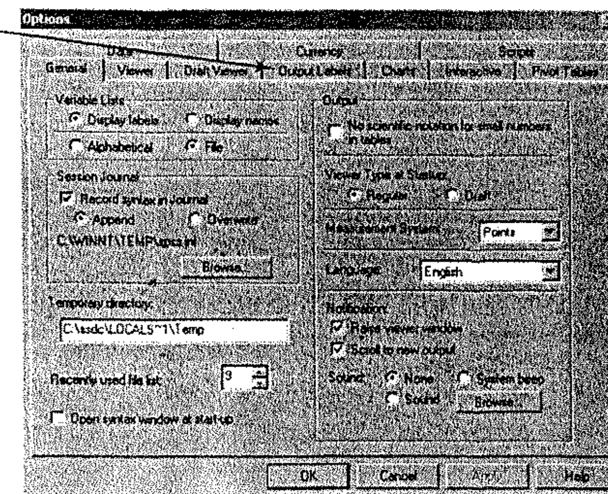
Pasul 1

Selectați „Edit" și apoi „Options...".



Pasul 2

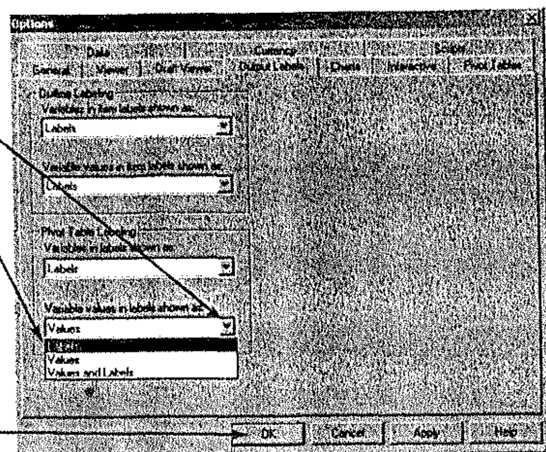
Selectați „Output Labels".



Pasul 3

Selectați butonul de sub „Variable values in labels shown as:" și „Labels" din meniul drop-down.

Apăsați „OK".



Etichetele sunt acum afișate

A treia coloană afișează procentul de frecvență pentru fiecare categorie, incluzând și valoarea lipsă, dar aici nu este nici una.

Deci 17 reprezintă 21,3% din totalul de 80 de oameni.

Occupation

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Nuns	17	21.3	21.3	21.3
	Nursery Teachers	3	3.8	3.8	25.0
	Television Presenters	23	28.8	28.8	53.8
	Students	20	25.0	25.0	78.8
	Other	17	21.3	21.3	100.0
	Total	80	100.0	100.0	

A patra coloană afișează procentul de frecvență excluzând valorile lipsă. Din moment ce nu avem nici o valoare lipsă, procentajul este același ca în coloana a treia.

A cincea coloană adună procentele în josul tabelului, astfel 25,0% din cazuri reprezintă fie categoria „Nuns", fie „Nursery Teachers".

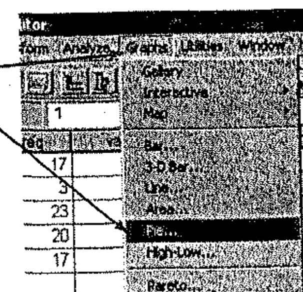
2.5. Raportarea rezultatelor

Numai etichetele categoriale, frecvențele și frecvențele procentuale trebuie să fie prezentate. În consecință acest tabel trebuie să fie simplificat dacă va fi prezentat. Dacă ocupația lipsește pentru unele cazuri, va trebui să decideți dacă veți prezenta procentajele incluzându-le sau excluzându-le pe acestea din urmă. Nu este necesar să prezentați ambele seturi de numere. De asemenea, omiteți termenul „Valid" din prima coloană, deoarece înțelesul său ar putea fi familiar numai pentru utilizatorii de SPSS.

2.6. Diagramă circulară pentru date categoriale

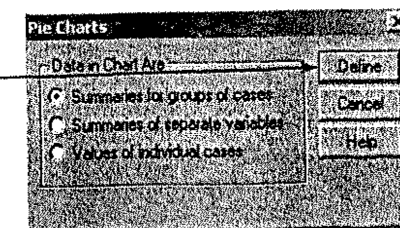
Pasul 1

Selectați „Graphs" și „Pie".



Pasul 2

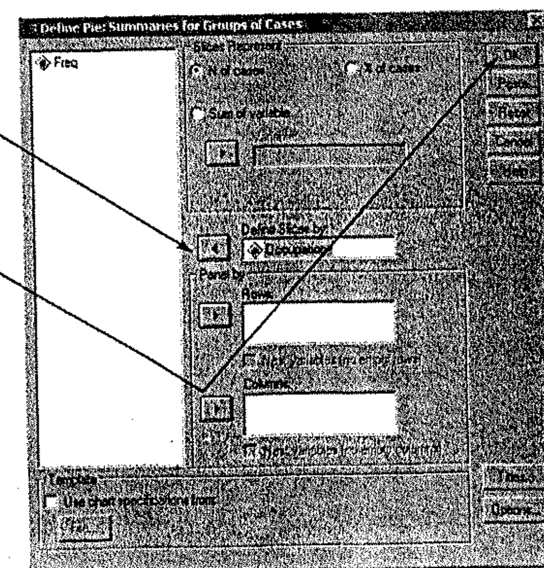
Selectați „Define".



Pasul 3

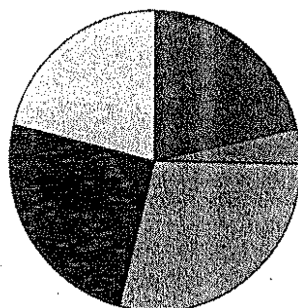
Selectați „Occupation" și apoi apăsați butonul ◀ de lângă „Define Slices by:" pentru a introduce variabila „Occupation".

Apăsați „OK".



Acesta este felul în care apare o diagramă circulară folosind opţiunile predefinite din SPSS. Sectoarele de cerc sunt codate cu diferite culori. Culorile nu au fost reproduse aici.

Caracteristicele din această diagramă pot fi modificate cu „Chart Editor". Vă vom arăta cum să etichetaţi fiecare sector pentru ca cititorul să nu fie nevoit să folosească codul culorilor, şi cum să schimbaţi culorile cu modele monocrome.



Occupation
Nuns
Nursery Teachers
Television Presenters
Students
Other

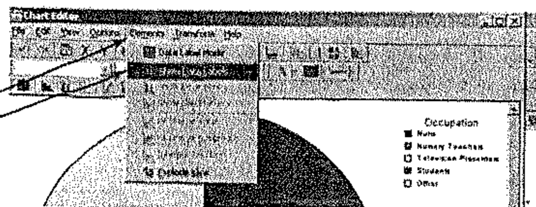
Cazurile comparate după frecvenţă

2.7. Adăugarea etichetelor unei diagrame circulare şi înlăturarea legendei şi a etichetei

Pasul 1

Daţi dublu clic oriunde în dreptunghiul care conţine diagrama, pentru a selecta „Chart Editor".

Selectaţi „Elements" şi „Show Data Labels".



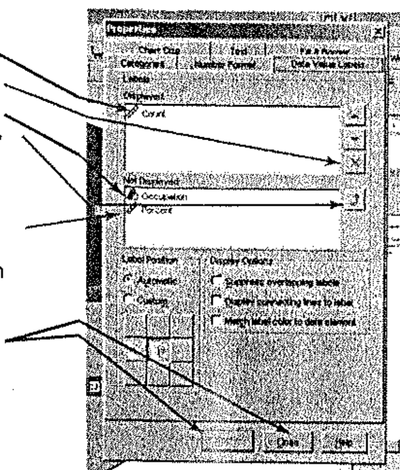
Pasul 2

Selectaţi „Count" şi „X"-ul roşu pentru a nu afişa.

Selectaţi „Occupation" şi săgeata verde curbată în sus, pentru a afişa denumirile ocupaţiilor.

Procedaţi asemănător pentru „Percent" pentru a afişa procentajul fiecărei ocupaţii în parte.

Selectaţi „Apply" şi „Close" dacă nu doriţi să faceţi vreo schimbare ulterioară a proprietăţilor diagramei circulare.



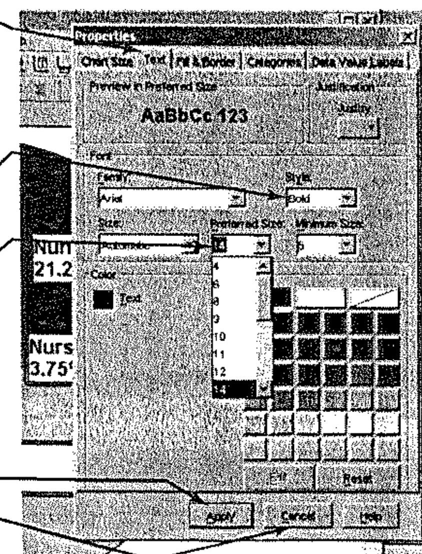
Pasul 3

Selectaţi „Text" dacă vreţi să schimbaţi formatul textului, cum ar fi caracterele îngroşate (Bold) şi dimensiunea acestora (Font Size).

Pentru a folosi caracterele îngroşate selectaţi butonul ▼ de lângă „Style" şi apoi „Bold" din meniul „dropdown".

Pentru a mări caracterele la dimensiunea 14, selectaţi butonul ▼ de lângă „Size" şi „14" din meniul care se va deschide mai jos.

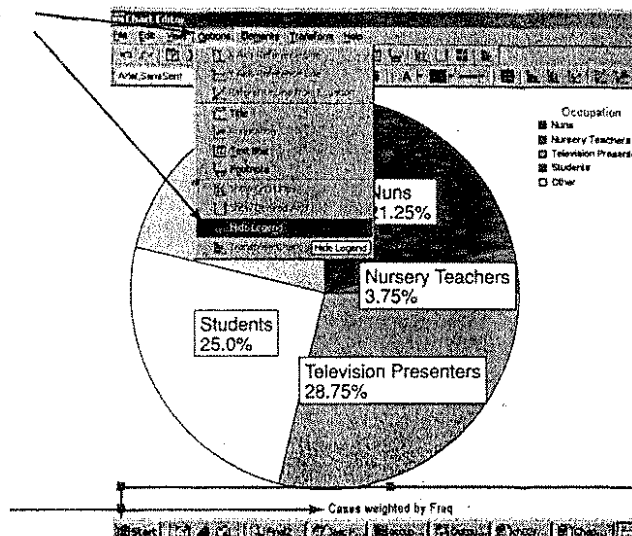
Dacă nu doriţi să faceţi vreo schimbare ulterioară a proprietăţilor diagramei circulare, selectaţi „Apply" şi „Close" (butonul „Cancel" se schimbă în „Close").



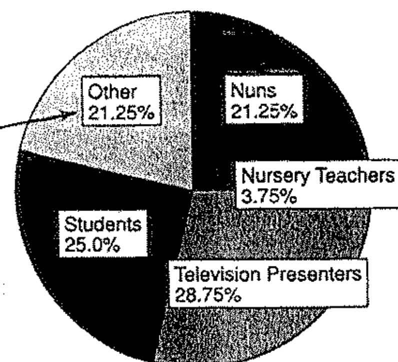
Pasul 4

Pentru a înlătura legenda, selectaţi „Options" şi „Hide Legend".

Pentru a înlătura „Cases weighted by Freq", daţi clic o dată pe etichetă, încă un clic, şi apoi stergeţi-l.



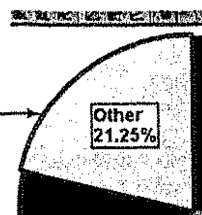
Aceasta este diagrama circulară cu sectoarele denumite şi frecvenţa procentuală afişată.



2.8. Schimbarea culorii unei diagrame circulare cu un model alb-negru

Pasul 1

În „Chart Editor", daţi clic pe sectorul de cerc pe care vreţi să-l schimbaţi, şi apoi încă un clic acolo unde marginea este o linie dublă.



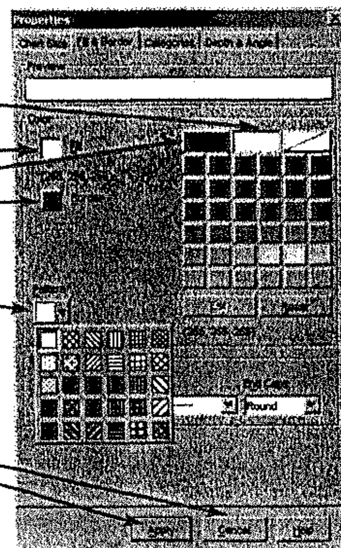
Pasul 2

Selectaţi „Edit" şi „Properties".

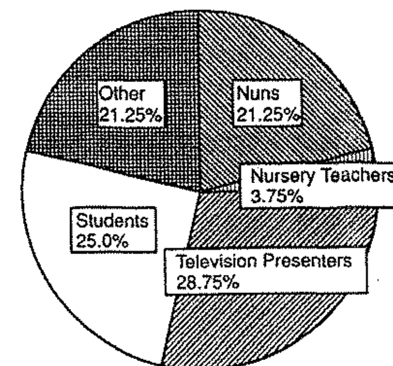
Selectaţi „Fill" şi culoarea albă.

Pentru a adăuga o margine neagră, selectaţi „Border" şi culoarea neagră.
Pentru a adăuga un model, selectaţi butonul ▾ de lângă „Pattern" şi modelul pe care îl doriţi.

Selectaţi „Apply" şi apoi „Close".
Folosiţi aceeaşi procedură pentru celelalte patru sectoare.



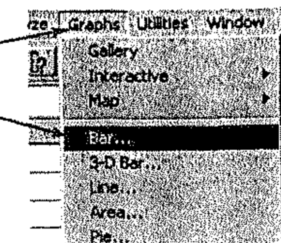
Aceasta este o diagramă circulară cu sectoare ce au modele alb-negru.



2.9. Diagramă cu bare pentru date categoriale

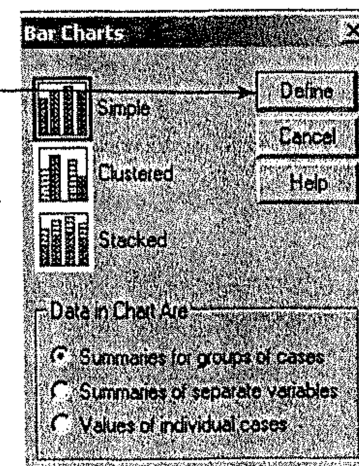
Pasul 1

Selectaţi „Graphs" şi „Bar...".



Pasul 2

Selectaţi „Define"; „Simple" este preselectat.

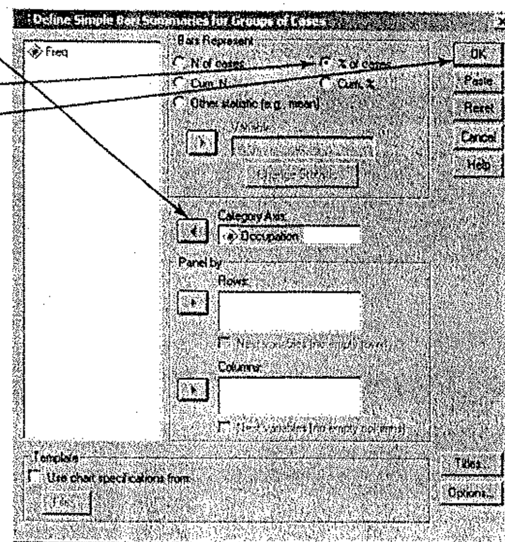


Pasul 3

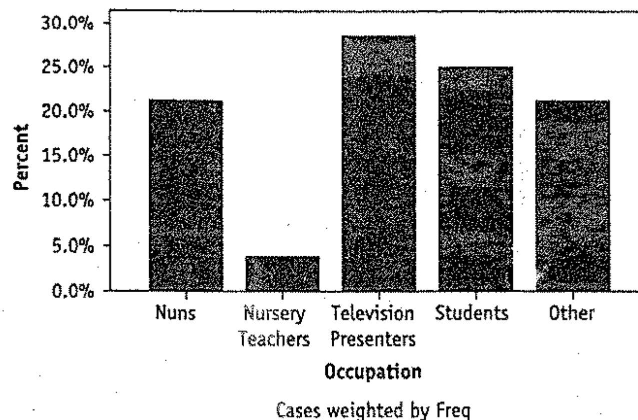
Selectați ocupația și apăsați butonul ► de lângă „Category Axis" pentru a introduce ocupația acolo.

Selectați „% of cases".

Apăsați „OK".



Aceasta este o diagramă cu bare. Poate fi editată cu „Chart Editor" dacă doriți.



2.10. Histograme

Vom ilustra procesul de realizare a unei histograme folosind datele din tabelul 2.2 care indică distribuția atitudinilor studenților față de statistică. Noi am etichetat această variabilă „Response".

Tabelul 2.2. *Distribuția răspunsurilor studenților la afirmația „Statistica este materia mea preferată la universitate"*

Response category	Value	Frequency
Strongly agree	1	17
Agree	2	14
Neither agree nor disagree	3	6
Disagree	4	2
Strongly disagree	5	1

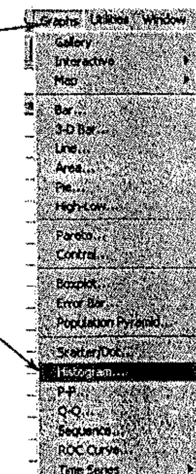
Pasul 1

În „Data Editor" introduceți datele, clasificați-le și etichetați-le așa cum s-a arătat la începutul acestui capitol.

	Response	Freq
1	1	17
2	2	14
3	3	6
4	4	2
5	5	1

Pasul 2

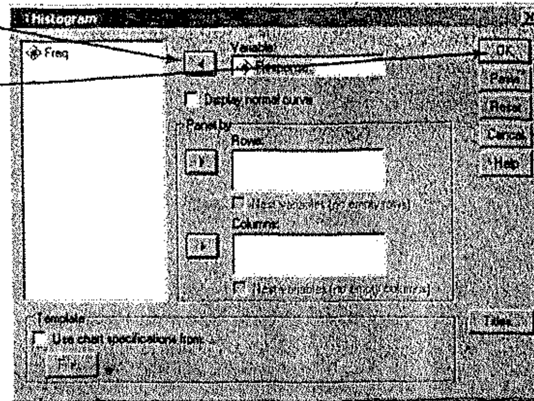
Selectați „Graphs" și „Histogram...".



Pasul 3

Selectați „Response" și apăsați butonul ◄ de lângă „Variable" pentru a duce variabila „Response" în acea casuță.

Apăsați „OK".

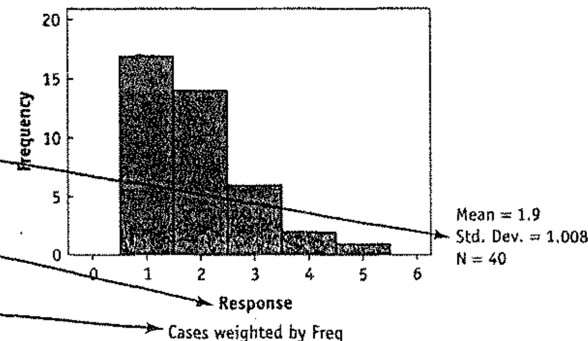


Aceasta este histograma, care poate fi editată cu „Chart Editor".

Pentru a înlătura statisticile, dați clic pe ele și ștergeți-le.

Pentru a schimba denumirile axelor dați clic pe ele și editați-le.

Pentru a înlătura „Cases weighed by Freq", dați clic pe etichetă, selectați-o și ștergeți-o.



3 Descrierea numerică a variabilelor
Medie, variație și dispersie

Rezumat

- În acest capitol vom descrie modul de lucru al unui număr de statistici, care rezumă și descriu caracteristicile esențiale pentru fiecare variabilă importantă. Tehnicile prezentate în acest capitol implică variabile individuale luate separat. Cu alte cuvinte, sunt tehnici statistice cu o singură variabilă sau univariate.
- Fiecare dintre ele generează un indice numeric pentru a descrie caracteristicile datelor.
- Cu excepția modului ce poate fi folosit pentru orice tip de date, celelalte tehnici sunt destinate datelor care au forma unor scoruri numerice.
- Indicatorul medie reprezintă media aritmetică a unui set de scoruri. Se obține prin însumarea valorilor și prin împărțirea rezultatului la numărul de valori.
- Modul este pur și simplu valoarea cu cea mai mare frecvență. Un set de valori pot dispune de mai mult de un mod, în cazul în care două sau mai multe valori apar cu frecvențe egale. Modul este valoarea variabilei care apare cel mai frecvent – nu este frecvența cu care apare cel mai frecvent scor.
- Mediana este valoarea din centrul distribuției, dacă variabilele numerice sunt ordonate după mărime, de la cea mai mică la cea mai mare. Din diverse motive, câteodată, mediana este o estimare a scorului mediu – de exemplu, când numărul scorurilor este egal, nu există o medie bine determinată.
- Procedurile descrise în acest capitol pot fi modificate rapid pentru a produce valorile varianței, boltirii și ale altor statistici descriptive.

Vom ilustra calculul mediei, al medianei și al modului folosind vârstele studenților de la universitate.

Tabelul 3.1. *Vârstele pentru 12 studenți*

18	21	23	18	19	19	19	33	18	19	19	20

3.1. Introducerea datelor

Pasul 1

În „Variable View" din „Data Editor" denumiți prima coloană „Age".

Îndepărtați cele două zecimale.

Name	Type	Width	Decimals
Age	Numeric	8	0

Pasul 2

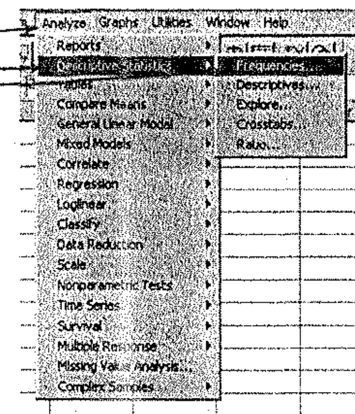
În „Data View" din „Data Editor" introduceți vârstele în prima coloană.

	Age
1	18
2	21
3	23
4	18
5	19
6	19
7	19
8	33
9	18
10	19
11	19
12	20

3.2. Efectuarea analizei

Pasul 3

Selectați „Analyze", „Descriptive Statistics" și „Frequencies...".

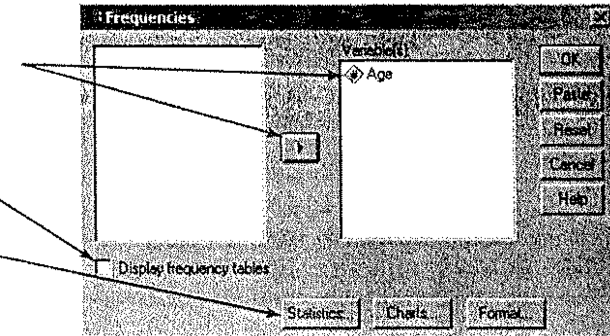


Pasul 4

Selectați „Age" și butonul ▶ pentru a-l introduce în lista de variabile.

Deselectați „Display frequency tables". Ignorați mesajul de avertizare.

Apoi dați clic pe „Statistics...".

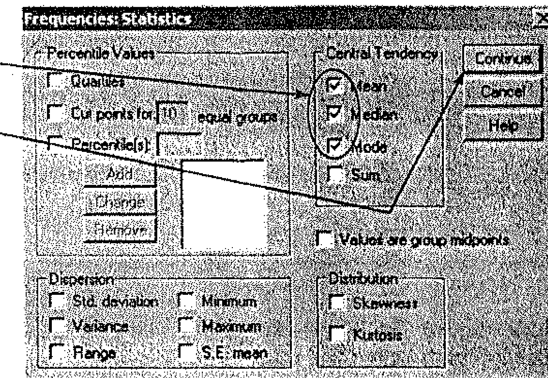


Pasul 5

Selectați „Mean", „Median", și „Mode".

Dați clic pe „Continue".

Apăsați „OK" din ecranul anterior, care reapare.



3.3. Interpretarea output-ului

Statistics

Age

N	Valid	12
	Missing	0
Mean		20.50
Median		19.00
Mode		19

Există 12 cazuri cu date valide pe care se bazează analiza.

Nu sunt date lipsă (0).

Media de vârstă media aritmetică = 20,5 ani.

Mediana de vârstă (vârsta persoanei aflate la mijlocul listei de vârste de la cea mai mică la cea mai mare) este de 19,00 ani.

Valoarea mod de vârstă (cea mai comună) este 19. Consultați următorul tabel de rezultate. Acesta furnizează frecvențele pentru fiecare valoare. O variabilă poate avea mai mult de un mod.

3.4. Raportarea output-ului

- Media, mediana și modul pot fi prezentate sub formă tabelară ca în tabelul 3.2.
- Pentru majoritatea datelor două zecimale sunt mai mult decât suficiente. Majoritatea măsurătorilor sunt aproximative și utilizarea mai multor zecimale tinde să dea o falsă senzație de precizie.
- Pentru mediană, este probabil mai clar dacă valoarea raportată este 19 în loc de 19,00. Totuși, dacă valorile de după virgulă sunt diferite de ,00, acest lucru ar trebui inclus în raport pentru că arată faptul că estimarea medianei nu corespunde nici unui scor.

Tabelul 3.2. *Media, mediana și modul pentru vârstă*

	Ages of students ($N = 12$)
Mean	20.50 years
Median	19 years
Mode	19 years

3.5. Alte caracteristici

Veți vedea din căsuțele de dialog de la pasul 5 că sunt multe alte valori statistice adiționale care pot fi calculate. Nu ar trebui să aveți mari dificultăți pentru a le obține, adaptând pașii prezentați anterior.

Centilele. Acestea indică punctele de separație pentru procentajele scorurilor. Așadar al 90-lea centil este valoarea numerică care separă cele 90% de valori de dedesubt, din punct de vedere al mărimii.

Cvartilele. Valorile distribuției care indică punctele de separare pentru cele mai mici 25%, cele mai mici 50% și cele mai mici 75% dintre scoruri.

Suma. Totalul scorurilor pentru o variabilă.

Skewness. Distribuțiile de frecvență nu sunt întotdeauna simetrice față de medie. Asimetria este un indice de asimetrie sau „inclinare" a distribuției scorurilor pentru o variabilă. Valoarea este pozitivă dacă valorile sunt asimetrice spre stânga, sau negativă dacă valorile sunt asimetrice către dreapta.

Kurtosis. Acesta este un index pentru cât de ascuțită sau turtită este distribuția scorurilor pentru o variabilă, comparativ cu distribuția normală. Va fi cu semnul „+" pentru curbe de frecvență ascuțite și cu semnul „-" pentru curbe turtite.

Abaterea standard (estimată). Este o evaluare a măsurii în care scorurile diferă în medie față de media scorurilor pentru o variabilă particulară. În SPSS abaterea standard este calculată ca o evaluare a abaterii standard a populației. Este un indice pentru variabilitatea scorurilor în jurul valorii medii a variabilei. Unii autori numesc acest indicator abatere standard.

Varianța (estimată). Aceasta este o evaluare a măsurii în care scorurilor variază în medie față de media scorurilor pentru variabila respectivă. Este pătratul abaterii standard și evident în strânsă legătură cu ea. Este de asemenea o estimare a varianței populației în SPSS. Unii autori numesc acest indicator varianță. Ca și abaterea standard, aceasta este un indice pentru variabilitatea scorurilor în jurul valorii medii a unei variabile, dar are, de asemenea, și alte utilizări în statistică. În special, este unitatea standard de măsurare în statistică.

Rang. Diferența numerică între cel mai mare și cel mai mic scor obținut pentru o variabilă. Este un singur număr.

Minim (scor). Valoarea celui mai mic scor al datelor pentru o variabilă particulară.

Maxim (scor). Valoarea celui mai mare scor al datelor pentru o variabilă particulară.

Eroarea standard (ES medie). Valoarea medie cu care mediile eșantioanelor extrase dintr-o populație, diferă față de media populației. Eroarea standard poate fi utilizată asemănător cu abaterea standard și varianța, ca un indice pentru variabilitatea scorurilor în jurul valorii medii a unei variabile.

4 Forme ale distribuţiilor scorurilor

Rezumat

- Este important de studiat forma distribuţiilor scorurilor pentru o variabilă. În mod ideal pentru majoritatea tehnicilor statistice, o distribuţie ar trebui să fie simetrică şi cu formă normală (formă de clopot).
- Unele tehnici statistice sunt foarte eficiente când distribuţiile variabilelor implicate sunt normal distribuite. Abaterile majore faţă de normalitate trebuie evitate, dar, pentru eşantioane de dimensiuni relativ mici, verificarea vizuală a diagramelor de frecvenţă este singurul mod practic de a le evita. Efectele eşecului de a îndeplini acest criteriu pot fi supraevaluate. Câteodată este posibilă transformarea statistică a unor scoruri pentru a aproxima o distribuţie normală, dar este în mare măsură o chestiune de încercare şi eroare, folosind, de exemplu, scale logaritmice.
- Totuşi, cercetătorii trebuie să fie precauţi faţă cu distribuţiile foarte asimetrice şi cele care conţin câteva scoruri neobişnuit de mici sau de mari (atipice). Histogramele, de exemplu, pot fi folosite pentru detectarea asimetriei şi a valorilor atipice.
- Cercetătorii trebuie să ia în calcul combinarea rangurilor scorurilor (în opoziţie cu tabelarea fiecărei valori individuale posibile) pentru a clarifica distribuţia datelor. Eşantioanele de dimensiuni mici, tipice multor lucrări în psihologie şi alte ştiinţe sociale, pot conduce la figuri sau diagrame risipite, în care tendinţele nu sunt clare.

Noi vom elabora un tabel de frecvenţă şi o histogramă cu scorurile obţinute la evaluarea extroversiunii pentru 50 de piloţi de linie aeriană, arătate în tabelul 4.1.

Tabelul 4.1. *Rezultatele obţinute la evaluarea extroversiunii pentru 50 de piloţi de linie aeriană*

3	5	5	4	4	5	5	3	5	2
1	2	5	3	2	1	2	3	3	3
4	2	5	5	4	2	4	5	1	5
5	3	3	4	1	4	2	5	1	2
3	2	5	4	2	1	2	3	4	1

4.1. Introducerea datelor

Pasul 1

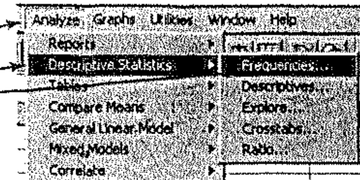
În „Variable View" din „Data Editor" denumiţi prima coloană „Extrav".

Înlăturaţi cele două zecimale.

Salvaţi aceste date ca un fişier pentru utilizare în capitolul 9.

Pasul 2

În „Data View" din „Data Editor" introduceţi scorurile la evaluarea extroversiunii în prima coloană.



4.2. Tabele de frecvenţă

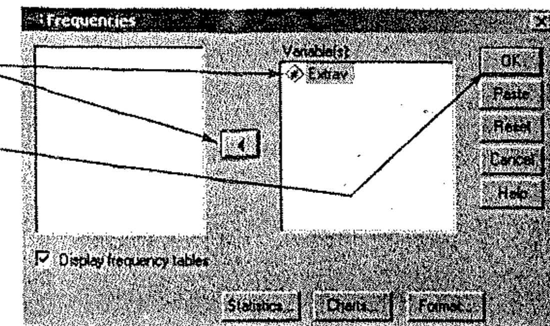
Pasul 3

Selectaţi „Analyze", „Descriptive Statistics" şi „Frequencies...".

Pasul 4

Selectaţi „Extrav" şi apoi apăsaţi butonul ◄ pentru a-l introduce în lista de variabile.

Apăsaţi „OK".



4.3. Interpretarea output-ului

Prima coloană arată cele cinci valori ale extroversiunii care sunt de la 1 la 5.

A doua coloană arată frecvenţa acestor valori. Sunt 7 cazuri în care valoarea este 1.

A treia coloană exprimă aceste frecvenţe ca un procent al numărului total incluzând şi datele care lipsesc. Din totalul cazurilor, 14% au valoarea 1.

Extrav

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	7	14.0	14.0	14.0
	2	11	22.0	22.0	36.0
	3	10	20.0	20.0	56.0
	4	9	18.0	18.0	74.0
	5	13	26.0	26.0	100.0
	Total	50	100.0	100.0	

A patra coloană exprimă aceste frecvenţe ca un procent al numărului total excluzând datele lipsă. Din moment ce nu sunt cazuri lipsă, procentajele sunt identice cu cele din coloana a treia.

A cincea coloană cumulează aceste procentaje în josul tabelului. Aşadar, 56% din cazuri au valoarea 3 sau valori mai mici.

4.4. Raportarea output-ului

Observaţi că în tabelul 4.2 am omis unele dintre detaliile neclare ale rezultatelor. Tabelele şi diagramele ar trebui să clarifice rezultatele.

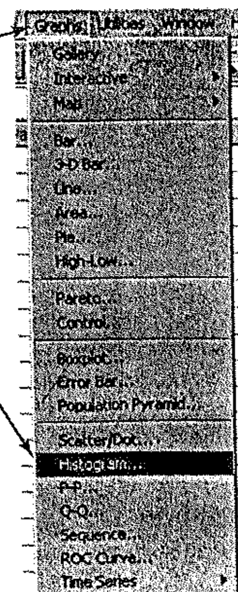
Tabelul 4.2. *Un mod de a raporta rezultatele din SPSS*

Extraversion score	Frequency	Percentage frequency	Cumulative percentage frequency
1	7	14.0	14.0
2	11	22.0	36.0
3	10	20.0	56.0
4	9	18.0	74.0
5	13	26.0	100.0

4.5. Histograme

Pasul 1

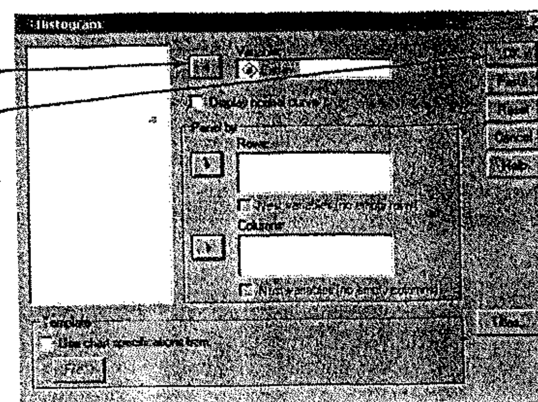
Selectaţi „Graphs" „Histogram...".



Pasul 2

Selectaţi „Extrav" şi apoi apăsaţi butonul ◄ pentru a o introduce în caseta de variabile.

Apăsaţi „OK".

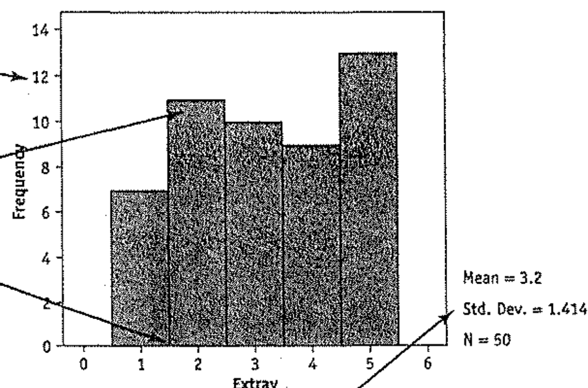


4.6. Interpretarea output-ului

Axa verticală arată frecvența cazurilor de la 0 la 14 din doi în doi.

Înălțimea barelor arată cât de frecvent apare fiecare scor la extroversiune.

Axa orizontală arată cele cinci valori diferite ale „Extrav" de la 1 la 5.



Media scorului la extroversiune este 3,2.

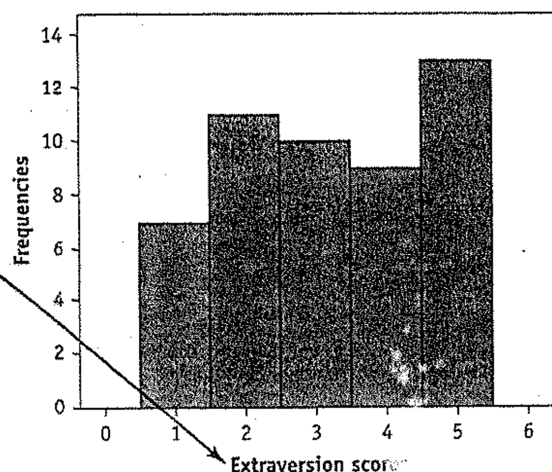
Abaterea standard este 1,414.

Numărul cazurilor este 50.

4.7. Raportarea histogramelor

Ar trebui să editați histograma înainte de a o raporta.

Schimbați „Frequency" în „Frequencies" și „Extrav" în „Extraversion scores". Aceste proceduri sunt descrise în capitolul 2.



5 Abaterea standard
Unitatea de măsură standard în statistică

Rezumat

- La bază, abaterea standard este un indice pentru cât de mult deviază (diferă) unele scoruri „în medie" față de media setului de scoruri din care acestea fac parte. Cu alte cuvinte, abaterea standard este un indice pentru variabilitatea scorurilor în jurul mediei acestora.
- Abaterea standard poate fi de asemenea folosită pentru a transforma scoruri pentru variabile foarte diferite în scoruri *z* (sau „standard"), care sunt ușor de comparat și însumat, din moment ce sunt plasate pe aceeași scală de măsurare. Această scală standard are întotdeauna o medie de zero și o abatere standard de 1,0 indiferent de variabilele în cauză.
- Scorurilele standard apar într-o varietate de feluri în statistică, deși nu totdeauna într-o formă evidentă. De exemplu, coeficientul beta din regresiile multiple se bazează pe notele standard. Scorurile *z* sunt ușor de procesat în SPSS, dar o înțelegere conceptuală a acestora este utilă în numeroase contexte.
- Abaterea standard, în sine, are întotdeauna o valoare pozitivă, *dar* cercetătorii menționează un număr de abateri standard peste medie („+" relativ la medie) sau pentru un număr de abateri standard sub medie („–" relativ la medie). Așadar, scorurile peste media eșantionului vor fi întotdeauna convertite la scoruri *z* (sau standard) cu un prefix pozitiv. Valorile sub medie vor fi întotdeauna convertite la scoruri *z* cu un prefix negativ.

Calcularea abatere standard și a scorurilor z este ilustrată cu nouă valori ale vârstei din tabelul 5.1.

Tabelul 5.1. *Date pentru calculul abaterii standard*

Age	20	25	19	35	19	17	15	30	27

5.1. Introducerea datelor

Pasul 1

În „Variable View" din „Data Editor" denumiți prima coloană „Age".

Înlăturați cele două zecimale.

Name	Type	Width	Decimals
Age	Numeric	8	0

Pasul 2

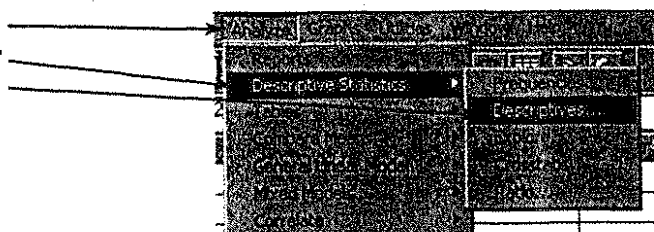
În „Data View" din „Data Editor" introduceți vârstele în prima coloană.

	Age
1	20
2	25
3	19
4	35
5	19
6	17
7	15
8	30
9	27
10	

5.2. Abaterea standard

Pasul 3

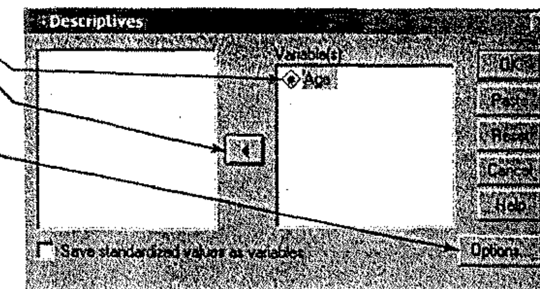
Selectați „Analyze", „Descriptive Statistics" și „Descriptives...".



Pasul 4

Selectați „Age" și apoi apăsați butonul ◄ pentru a o introduce în lista de variabile, „Variable(s):".

Selectați „Options...".

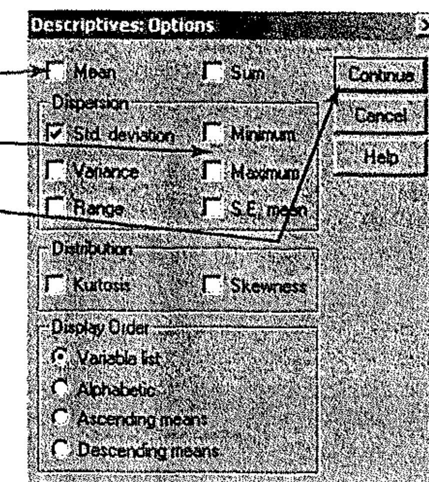


Pasul 5

Deselectați „Mean", „Minimum" și „Maximum".

Selectați „Continue".

Selectați „OK" din ecranul precedent, care reapare.



5.3. Interpretarea output-ului

Numărul de cazuri este 9.

Abaterea standard pentru „Age" este 6,652.

Descriptive Statistics

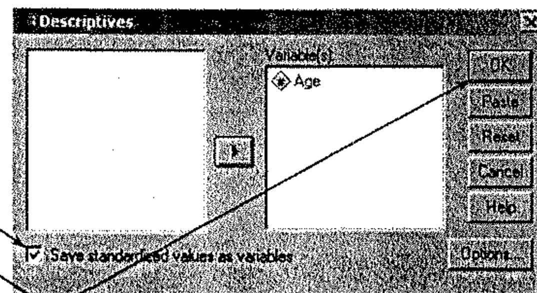
	N	Std. Deviation
Age	9	6.652
Valid N (listwise)	9	

5.4. Scorurile *z*

Pasul 1

La pasul 4 selectaţi „Save standardized values as variables".

Apăsaţi „OK".



Pasul 2

Scorurile *z* sau standard sunt în a doua coloană din „Data View" în „Data Editor" şi sunt numite „ZAge".

	Age	ZAge
1	20	-.45099
2	25	.30066
3	19	-.60132
4	35	1.80395
5	19	-.60132
6	17	-.90198
7	15	-1.20263
8	30	1.05230
9	27	.60132

5.5. Raportarea output-ului

Abaterea standard pentru o variabilă este uşor de menţionat în textul raportului: „A fost determinată abaterea standard a vârstei ca fiind 6,65 ani (N = 9)".

Însă este mai probabil că veţi vrea să înregistraţi abaterea standard împreună cu alte statistici, cum ar fi media, rangul, cum este ilustrat în tabelul 5.2. Veţi dori probabil să includeţi aceste statistici pentru alte variabile numerice pentru care aveţi date.

Tabelul 5.2. *Dimensiunea eşantionului, media, rangul şi abaterea standard pentru vârstă, IQ şi fluenţă verbală*

	N	Mean	Range	Standard deviation
Age	9	23.00	20.00	6.65
IQ	9	122.17	17.42	14.38
Verbal fluency	9	18.23	4.91	2.36

5.6. Alte caracteristici

Meniul „Descriptives..." conţine un număr de calcule statistice care pot fi selectate:

Mean
Sum
Standard deviation (estimate)
Range
Minimum
Maximum
Kurtosis
Skewness

Aceste concepte statistice sunt explicate pe scurt la sfârşitul capitolului 3.

6 Relațiile dintre două sau mai multe variabile
Diagrame şi tabele

Rezumat

- O mare parte a cercetării studiază relațiile dintre două sau mai multe variabile. Procedurile statistice univariate (o singură variabilă) descrise până acum pot fi utilizate pentru analiza, practic, a oricăror date. Totuşi, multe dintre întrebările cercetării necesită, de asemenea, ca relațiile şi corelațiile dintre diferite variabile să fie studiate.
- La fel ca pentru statistica univariată, o analiză statistică bivariată completă a datelor necesită studierea tendințelor fundamentale ale datelor utilizând tabele şi diagrame. Metodele de prezentare a relațiilor bivariate includ crearea tabelelor de asociere, histograme compuse (diagrame cu bare grupate) şi diagrame scatter. Multe dintre acestea sunt deja familiare, dar totuşi pot genera dificultăți.
- Trebuie să fiți siguri că tabelele şi diagramele obținute sunt folositoare şi clare. În particular, asigurați-vă că datele pentru tabelele de asociere şi histogramele compuse (diagrame cu bare grupate) conțin numai un număr mic de valori diferite ale datelor. Dacă nu, SPPS-ul va produce tabele masive, grafice şi diagrame dense şi ilizibile.
- Etichetarea tabelelor şi a diagramelor în întregime este o condiție de bază, împreună cu un titlu clar.
- Pe de altă parte, diagramele scatter funcționează bine şi când aveți multe valori diferite ale variabilelor.
- Tipul tabelului sau al diagramei care este cel mai eficient pentru comunicarea relațiilor dintre date depinde foarte mult de tipurile datelor implicate. Două variabile numerice, în general, vor fi cel mai eficient prezentate sub forma unei diagrame scatter, în timp ce două variabile nominale sunt în general prezentate sub forma unui tabel de asociere.
- Amintiți-vă că pentru prezentarea eficientă a statisticii descriptive fundamentale cercetătorii trebuie să planifice cu grijă tabelele şi diagramele. Adeseori cercetătorii vor trebui să „altereze" lucrurile pentru a transforma tabelele şi graficele finale în forme eficiente de comunicare. Faptul că acest lucru poate fi efectuat rapid este unul dintre avantajele utilizării SPSS-ului. Este recomandat să fie evitată tentația de a copia pur şi simplu tabelele şi diagramele SPSS-ului în raport. Nu este aproape nici o prezentare a rezultatelor în SPSS care să nu poată fi îmbunătățită şi clarificată. Opțiunile de bază din SPSS nu sunt întotdeauna ideale pentru acest aspect. Adeseori, procedurile de editare disponibile pentru tabelele din SPSS pot îmbunătăți enorm lucrurile.
- Realizarea de tabele, ca acelea discutate în acest capitol, este una dintre sarcinile dificile în SPSS.

Vom ilustra elaborarea unui tabel de asociere şi a unei diagrame cu bare compusă, cu datele din tabelul 6.1 (*ISP*, tabelul 6.4). Acesta prezintă numărul de bărbați şi femei dintr-un studiu care au fost sau nu spitalizați anterior. Dacă datele au fost deja introduse în SPSS, atunci paşii de la 1 la 5 pot fi ignorați.

Tabelul 6.1. *Tabel de asociere între variabilele gen şi spitalizare*

	Male	Female
Previously hospitalised	$f = 20$	$f = 25$
Not previously hospitalised	$f = 30$	$f = 14$

6.1. Introducerea datelor

Pasul 1

Cea mai rapidă cale de a introduce datele în tabelul 6.1 este de a crea cele patru celule sub forma a patru rânduri.

Pentru aceasta avem nevoie de trei variabile.

În „Variable View" din „Data Editor" denumiți prima coloană „Hospitalisation", a doua coloană „Gender" şi a treia coloană „Freq".

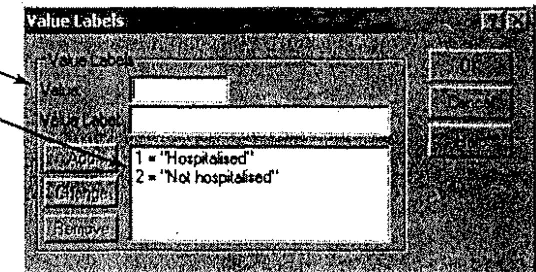
Înlăturați cele două zecimale.

	Name	Type	Width	Decimals	Label	Values
1	Hospitalisat	Numeric	8	0	Hospitalisation	{1, Hospitalise
2	Gender	Numeric	8	0	Gender	{1, Males}...
3	Freq	Numeric	8	0	Freq	None

Pasul 2

Etichetați cele două valori ale „Hospitalisation", (1 = „Hospitalised", 2 = „Not hospitalised") şi „Gender", Sex (1 = „Male",2 = „Female").

Paşii acestei proceduri au fost explicați în capitolul 1.



Pasul **3**

Introduceți aceste numere în „Data View" din „Data Editor".

Primul rând se referă la bărbații care au fost spitalizați; care sunt în număr de 20.

Al doilea rând se referă la femeile care au fost spitalizate; care sunt în număr de 25.

Al treilea rând se referă la bărbații nespitalizați; care sunt în număr de 30.

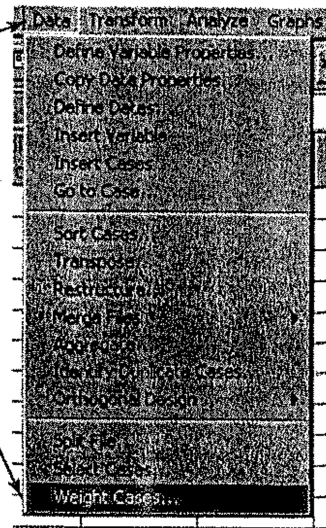
Al patrulea rând se referă la femeile nespitalizate; care sunt în număr de 14.

	Hospitalisation	Gender	Freq
1	1	1	20
2	1	2	25
3	2	1	30
4	2	2	14

6.2. Ponderarea datelor

Pasul **4**

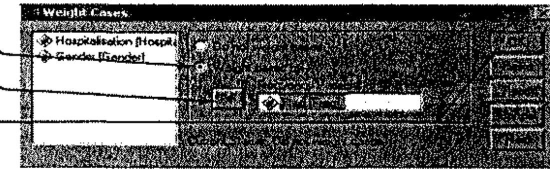
Pentru a pondera datele cu scopul ca cele patru celule să conțină numărul de cazuri corespunzător, selectați „Data" și „Weight Cases...".



Pasul **5**

Selectați „Freq", „Weight cases by" și apoi apăsați butonul ◄ pentru transfer în căsuța variabilei „Frequency Variable:".

Apăsați „OK".



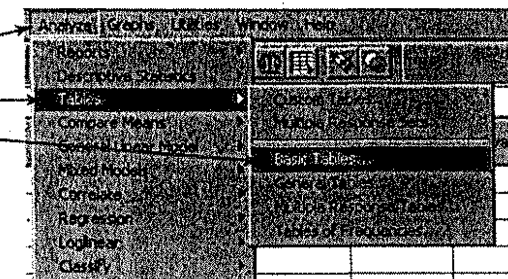
Cazurile sunt acum ponderate în funcție de frecvențe, așa cum este arătat de mesajul „Weight On" din colțul din dreapta jos al ferestrei „Data Editor".



6.3. Crearea tabelelor de asociere

Pasul **6**

Selectați „Analyze", „Tabels" și „Basic Tables...".

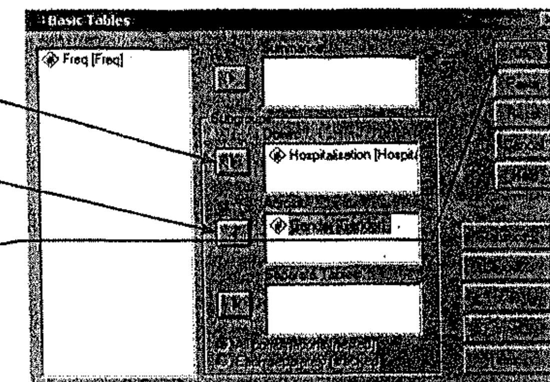


Pasul **7**

Pentru a pune „Hospitalisation" pe rândurile tabelului, selectați-o și apoi apăsați butonul ► de lângă eticheta „Down:".

Pentru a pune „Gender" pe coloanele tabelului, selectați-o și apoi apăsați butonul ◄ de lângă eticheta „Across:".

Apăsați „OK".



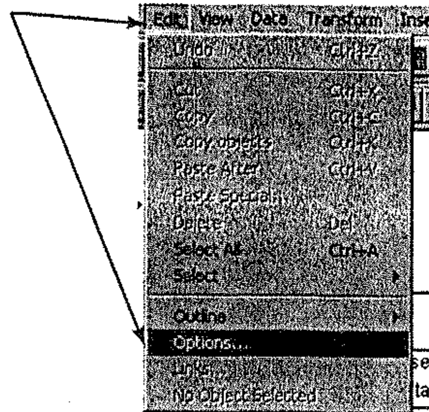
Pasul 8

Tabelul din output este structurat ca tabelul 6.1, cu excepția că cele două valori pentru fiecare variabilă nu sunt etichetate. Dacă este necesar etichetați aceste valori folosind pașii 9, 10 și 11 și reluați analiza.

		Gender	
		1	2
Hospitalisation	1	20	25
	2	30	14

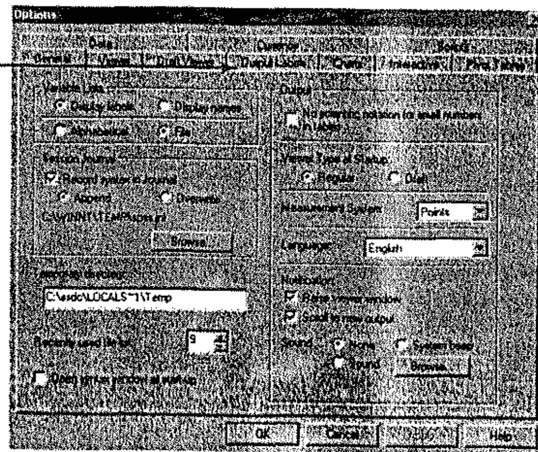
Pasul 9

Selectați „Edit" și „Options...".



Pasul 10

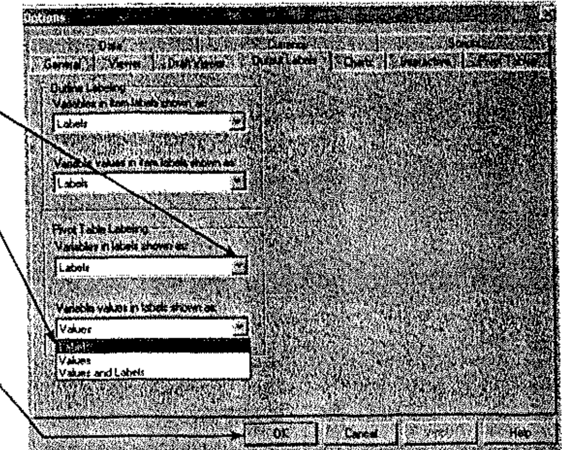
Selectați „Output Labels".



Pasul 11

Selectați butonul ▼ de sub eticheta „Variable values in labels shown as:" și „Labels" din meniul drop-down.

Apăsați „OK".



Pasul 12

Acesta este tabelul cu valorile etichetate.

		Gender	
		Males	Females
Hospitalisation	Hospitalised	20	25
	Not hospitalised	30	14

6.4. Afișarea frecvențelor ca procentaj din numărul total

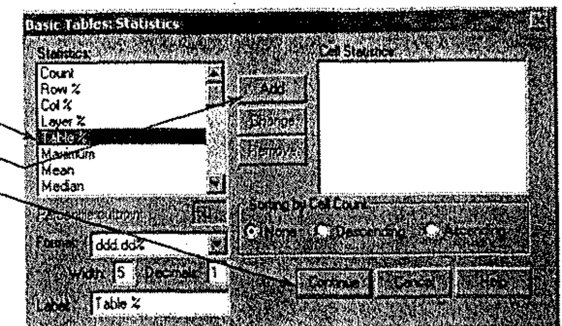
La pasul 7 selectați „Statistics...".

Selectați „Tables%".

Apoi selectați „Add".

Apăsați „Continue".

Apăsați „OK".



Dacă veți aduna procentajele din cele patru celule, suma lor este 100.

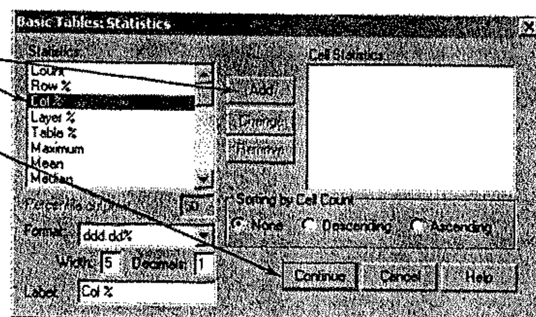
		Gender	
		1	2
		Table %	Table %
Hospitalisation	1	22.5%	28.1%
	2	33.7%	15.7%

6.5. Afişarea frecvenţelor ca procentaj din suma coloanei

Selectați „Col%" și „Add".

Apăsați „Continue".

Apăsați „OK" din ecranul anterior, care va reapărea.



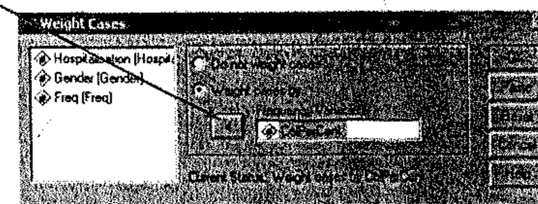
Dacă adunați procentajele de pe fiecare coloană, ele vor însuma 100.

		Gender	
		1	2
		Col %	Col %
Hospitalisation	1	40.0%	64.1%
	2	60.0%	35.9%

6.6. Diagrama procentuală compusă cu bare (suprapuse)

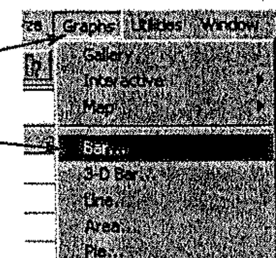
Pasul 1

Pentru a obține o diagramă procentuală cu bare compuse (suprapuse) în care barele să reprezinte 100%, va trebui să introduceți cifrele de procentaj (coloană numită „ColPerCent") pentru cele două bare și să le ponderați.



Pasul 2

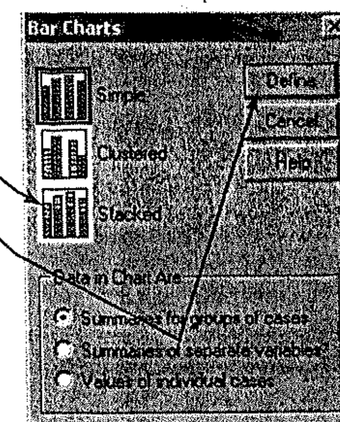
Selectați „Graphs" și „Bar...".



Pasul 3

Selectați „Stacked"

Selectați „Define".

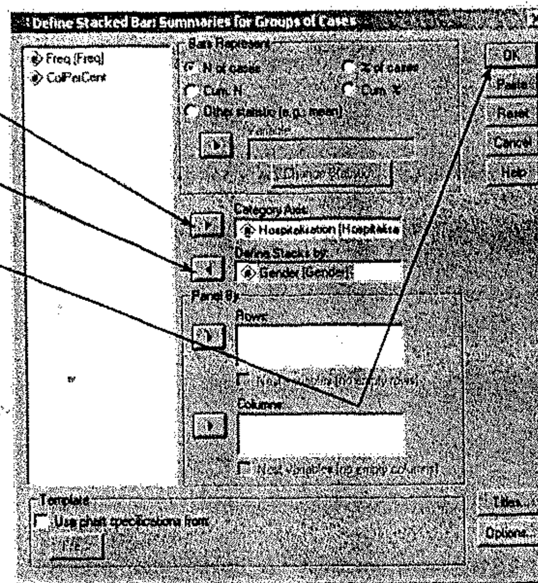


Pasul 4

Selectați „Hospitalisation" și apoi butonul ▶ de lângă eticheta „Category Axis:".

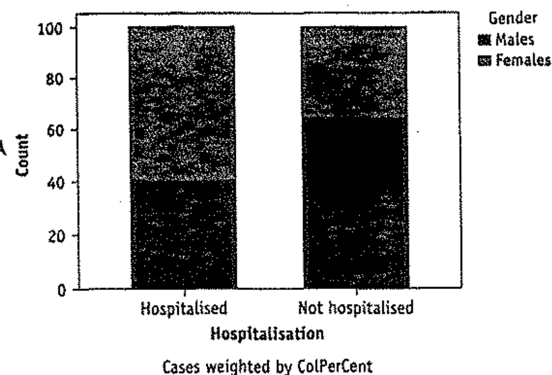
Selectați „Gender" și apoi butonul ◀ de lângă eticheta „Define Stacks by:".

Apăsați „OK".



Pasul 5

„Count" se referă la „Percent", dar poate fi schimbat cu „Chart Editor" (vezi capitolul 2).

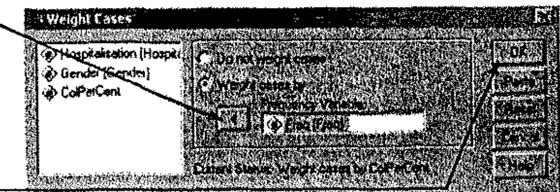


6.7. Histogramă compusă (diagrame cu bare grupate)

Pasul 1

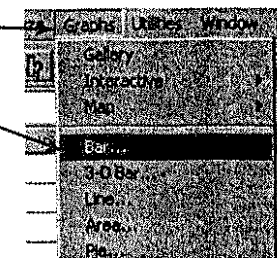
Ponderați cazurile după „Freq" în loc de „ColPerCent", selectând „Data", „Weight Cases...", „Freq", „Weight cases by" și apăsând butonul ◀ pentru a introduce frecvența în căsuța variabilei „Frequency Variable:".

Apăsați „OK".



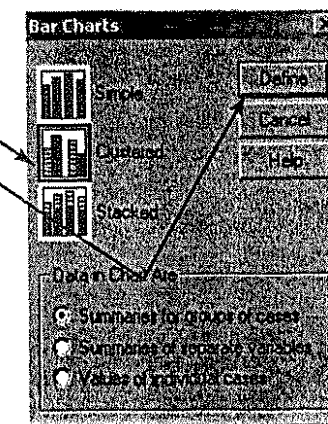
Pasul 2

Selectați „Graphs", apoi „Bar...".



Pasul 3

Selectați „Clustered" și apoi „Define".



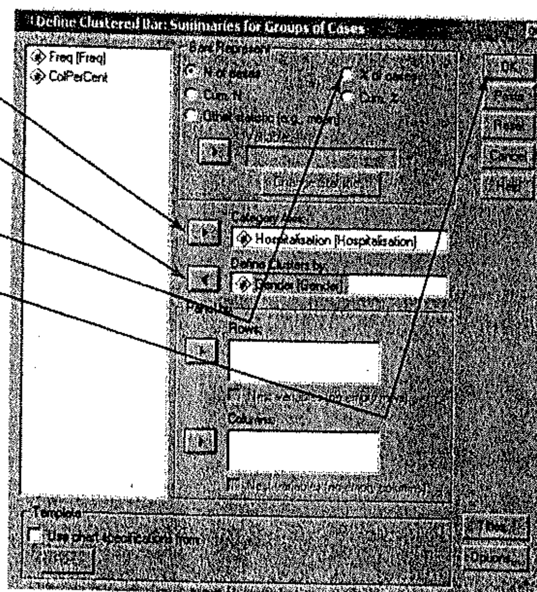
Pasul 4

Selectaţi „Hospitalisation" şi apoi apăsaţi butonul ▶ de lângă eticheta „Category Axis:".

Selectaţi „Gender" şi apoi butonul ▶ de lângă eticheta „Define Stacks by:".

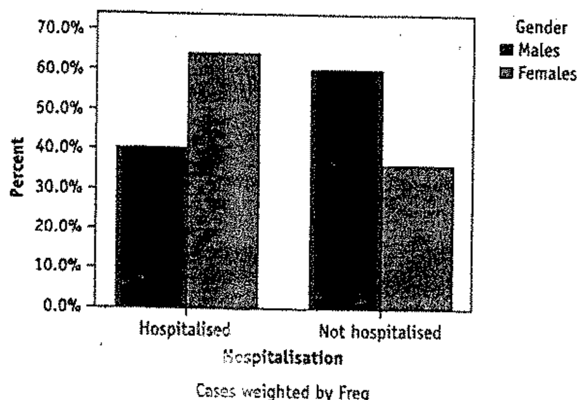
Selectaţi „% of cases".

Apăsaţi „OK".



Pasul 5

Aceasta este diagrama cu bare realizată.



7 Coeficienţi de corelaţie

Coeficienţii de corelaţie Pearson şi Spearman

Rezumat

- Există mai mulţi coeficienţi de corelaţie. În general, cel mai comun şi cel mai folositor este de departe coeficientul de corelaţie Pearson. Corelaţia biserial phi şi coeficientul de corelaţie Spearman sunt doar variante ale sale.
- Este o practică utilă să desenaţi o diagramă scatter pentru că aceasta prezintă datele incluse într-un coeficient de corelaţie. Veţi avea nu numai o reprezentare vizuală a relaţiei, ci şi un ajutor pentru identificarea unor serii de probleme cum ar fi relaţiile curbe şi prezenţa valorilor atipice. Coeficientul de corelaţie Pearson presupune o relaţie în linie dreaptă între două variabile. Este înşelător dacă între două variabile există o relaţie curbă. Valorile atipice sunt scorurile extreme şi neobişnuite care distorsionează mărimea coeficientului de corelaţie. Printre remedii este inclusă examinarea relaţiei omiţând valorile atipice. În mod alternativ, un coeficient de corelaţie Spearman este mai puţin afectat de valorile atipice, deci se poate examina şi mărimea coeficientului Spearman pentru aceleaşi date.
- Un coeficient de corelaţie este o măsură numerică sau un indice al gradului de asociere între două seturi de scoruri. Intervalul de valori este de la maxim +1,00, prin 0,00, până la –1,00.
- Semnul „+" indică o corelaţie pozitivă – anume că scorurile unei variabile se măresc odată cu creşterea scorurilor celeilalte variabile. Semnul „–" indică o corelaţie negativă – anume că, în timp ce scorurile unei variabile cresc, scorurile pentru cealaltă variabilă descresc.
- O corelaţie de 1,00 indică o asociere perfectă între cele două variabile. Cu alte cuvinte, o diagramă scatter a celor două variabile va arăta că toate punctele sunt conţinute de o singură dreaptă. O valoare de 0,00 indică faptul că toate punctele din diagrama scatter sunt dispersate aleatoriu în jurul oricărei drepte desenate pe această diagramă a datelor sau sunt aranjate într-o manieră curbilinie. O corelaţie de –0,5 ar indica o relaţie moderat negativă între cele două variabile.
- Coeficientul Spearman este coeficientul de corelaţie Pearson aplicat unor scopuri după ce acestea au fost ordonate crescător pentru cele două variabile în parte. Este folosit atunci când premisele coeficientului de corelaţie Pearson nu sunt îndeplinite de către date – această situaţie se petrece în special în momentul în care scorurile pentru o variabilă au o asimetrie pronunţată.

- Din moment ce coeficienţii de corelaţie sunt de obicei bazaţi pe eşantioane de date, este obişnuită incluziunea explicaţiei privind importanţa statistică a coeficientului de corelaţie. Importanţa statistică este o declaraţie a probabilităţii de a obţine un coeficient de corelaţie particular pentru un eşantion de date dacă nu există nici o corelaţie (corelaţie de 0,00) în cazul populaţiei din care a fost extras eşantionul. SPSS-ul poate oferi semnificaţii statistice sub forma unei valori exacte sau a unuia dintre nivele de semnificaţie convenţionale critice (de exemplu, 0,05 şi 0,01).

Vom ilustra calculul corelaţiei Pearson, o diagramă scatter şi coeficientul de corelaţie Spearman pentru datele din tabelul 7.1, care prezintă scoruri ale abilităţilor muzicale şi matematice pentru 10 copii.

Tabelul 7.1. *Scopurile abilităţilor muzicale şi matematice pentru 10 copii*

Music score	Mathematics score
2	8
6	3
4	9
5	7
7	2
7	3
2	9
3	8
5	6
4	7

7.1. Introducerea datelor

Pasul 1

În „Variable View" din „Data Editor" denumiţi primul rând „Music" şi al doilea „Math".

Înlăturaţi cele două zecimale schimbând cifra prezentă aici în zero.

	Name	Type	Width	Decimals
1	Music	Numeric	8	0
2	Maths	Numeric	8	0

Pasul 2

În „Data View" din „Data Editor" introduceţi datele pentru muzică în prima coloană şi cele pentru matematică în cea de-a doua coloană.

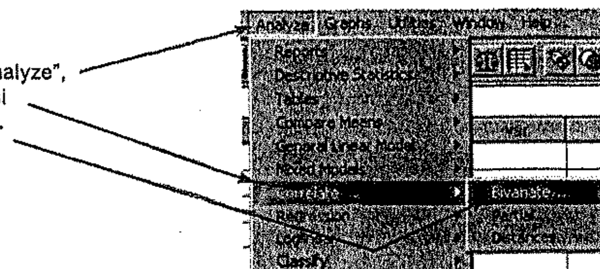
Salvaţi datele ca un fişier pentru utilizare în capitolul 8.

	Music	Maths
1	2	8
2	6	3
3	4	9
4	5	7
5	7	2
6	7	3
7	2	9
8	3	8
9	5	6
10	4	7

7.2. Corelaţia Pearson

Pasul 1

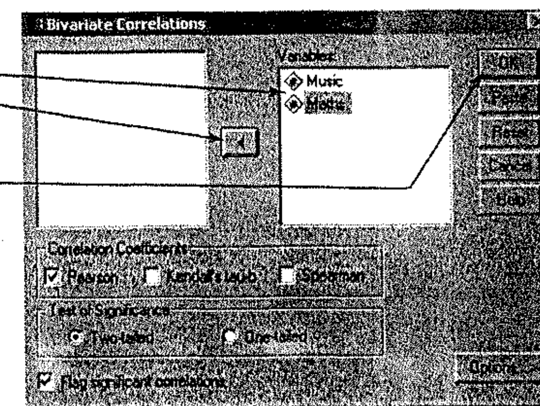
Selectaţi „Analyze", „Correlate" şi „Bivariate...".



Pasul 2

Selectaţi „Music" şi „Math" împreună sau separat şi apăsaţi butonul ◄ pentru a le introduce în lista de variabile aşa cum este arătat aici.

Apăsaţi „OK".



7.3. Interpretarea output-ului

Corelaţia dintre „Math" şi „Music" este –0,900. La un test de semnificaţie, two-tailed la nivelul de probabilitate de 0,001 sau mai mic, corelaţia este statistic semnificativă.

Numărul cazurilor pe care este bazată corelaţia este 10. Informaţia apare şi în această celulă.

Correlations

		Music	Maths
Music	Pearson Correlation	1	-.900**
	Sig. (2-tailed)	.	.000
	N	10	10
Maths	Pearson Correlation	-.900**	1
	Sig. (2-tailed)	.000	.
	N	10	10

**. Correlation is significant at the 0.01 level

7.4. Raportarea output-ului

- Corelaţia dintre abilităţile muzicale şi cele matematice este –0,900. Se obişnuieşte să se rotunjească corelaţia cu două zecimale, deci aceasta va deveni –0,90. Acesta este un rezultat mai mult decât precis pentru majoritatea măsurătorilor psihologice.
- Nivelul de semnificaţie exact, cu trei zecimale, este 0,000. Înseamnă că nivelul de semnificaţie este mai mic decât 0,001. Vă sugerăm să nu folosiţi un şir de zerouri, pentru că derutează cititorii. Întotdeauna schimbaţi ultimul zero în 1. Aceasta înseamnă că nivelul de semnificaţie poate fi raportat ca fiind $p < 0,001$.
- Se obişnuieşte să fie prezentate gradele de libertate (*DF*) în locul numărului de cazuri, când se prezintă corelaţiile. Gradele de libertate sunt egale cu numărul de cazuri minus 2, adică 8 pentru această corelaţie. Nu este nimic greşit în a raporta numărul de cazuri în schimb.
- Într-un raport, noi am scrie: „Există o relaţie negativă semnificativă între abilităţile muzicale şi cele matematice ($r = -0,90$, $DF = 8$, $p < 0,001$). Copiii cu nivel ridicat al abilităţi muzicale au un nivel scăzut al abilităţii matematice". Semnificaţia coeficientului de corelaţie este discutată în detaliu în acest manual (capitolul 10).

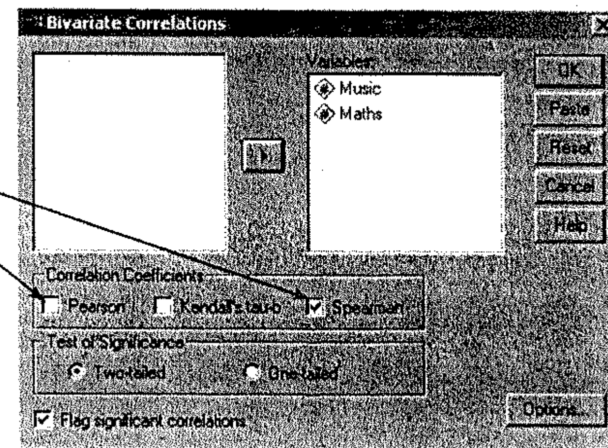
7.5. Coeficientul de corelaţie Spearman

Pasul 1

Identic corelaţiei Pearson, selectaţi „Analyze", „Correlate", „Bivariate" şi variabilele pe care vreţi să le corelaţi.

Selectaţi „Spearman".

Dacă nu doriţi Pearson, deselectaţi-l.



7.6. Interpretarea output-ului

Coeficientul de corelaţie Spearman între „Math" şi „Music" este –0,894.

Nivelul de semnificaţie, luând în calcul două ipoteze, este 0,001 sau mai mic, deci corelaţia este statistic semnificativă.

Numărul cazurilor este 10. Informaţia este oferită şi în această celulă.

Correlations

			Music	Maths
Spearman's rho	Music	Correlation Coefficient	1.000	-.894**
		Sig. (2-tailed)	.	.000
		N	10	10
	Maths	Correlation Coefficient	-.894**	1.000
		Sig. (2-tailed)	.000	.
		N	10	10

**. Correlation is significant at the 0.01 level (2-tailed).

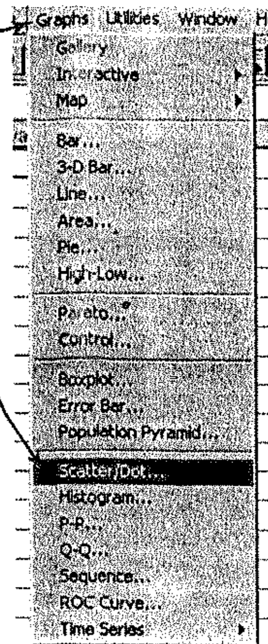
7.7. Raportarea output-ului

- Corelaţia raportată cu două zecimale este –0,89.
- Probabilitatea de a obţine această corelaţie din întâmplare este mai mică de 0,001 ($p < 0,001$).
- Noi am raporta astfel: „Există o corelaţie negativă semnificativă între abilităţile muzicale şi cele matematice (Coeficientul Spearman = $-0,89$, $DF = 8$, $p < 0,001$). Cei cu nivel mai ridicat al abilităţii muzicale au abilităţi matematice scăzute şi viceversa".

7.8. Diagrama scatter

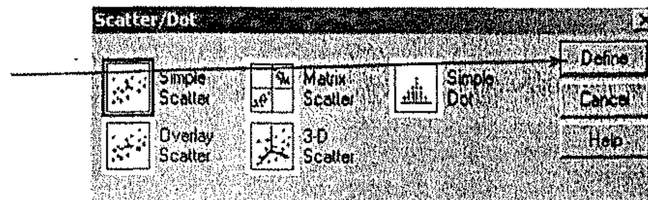
Pasul 1

Selectați „Grap
„Scatter/Dot...”.



Pasul 2

Selectați „Define” din moment ce „Simple” este deja selectat. Se numește opțiunea predefinită.

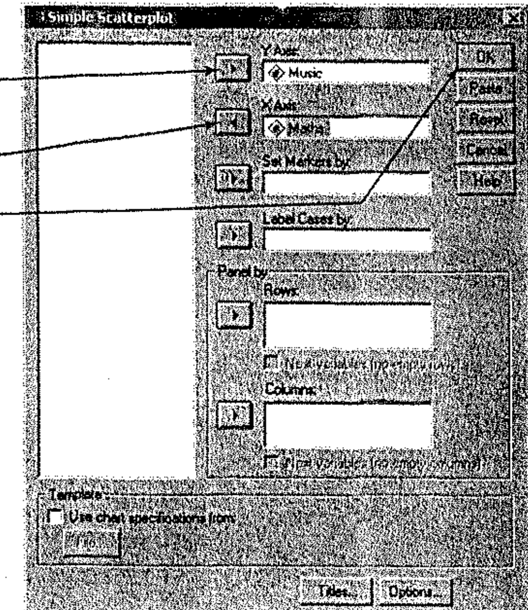


Pasul 3

Pentru a avea variabila „Music” ca axă verticală, selectați-o și apoi apăsați butonul ▶ de lângă căsuța „Y Axis”.

Pentru a avea variabila „Math” ca axă orizontală, selectați-o și apoi apăsați butonul ◀ de lângă căsuța „X Axis”.

Apăsați „OK”.



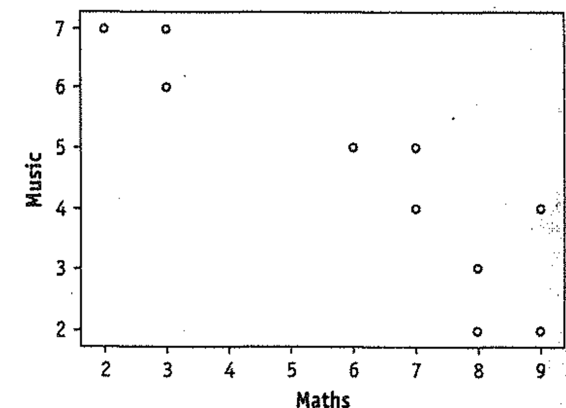
7.9. Interpretarea output-ului

În diagrama scatter, răspândirea punctelor este relativ îngustă, ceea ce indică o corelație mare.

Forma împrăștierii rezultatelor este relativ în linie dreaptă, indicând mai degrabă o relație în linie dreaptă decât o relație curbilinie.

Linia este de la colțul stânga sus până în dreapta jos, ceea ce indică o corelație negativă.

Dacă relația este curbilinie, atunci corelațiile Pearson și Spearman pot fi înșelătoare.



7.10. Raportarea output-ului

- Nu ar trebui să raportaţi niciodată un coeficient de corelaţie fără a examina diagrama scatter, din cauza unor probleme cum ar fi relaţiile curbilinii sau valorile atipice (*ISP*, capitolul 7).
- Într-un proiect studenţesc ar trebui să fie întotdeauna posibilă includerea de grafice de acest fel. Din păcate, articolele de jurnal sau cărţile tind să aibă restricţii referitoare la numărul de grafice inclus, din cauza spaţiului pe care îl ocupă şi a impactului asupra costului.
- Am putea scrie despre diagrama scatter : „A fost examinată o diagramă scatter pentru relaţia dintre abilităţile muzicale şi matematice. Nu există nici o dovadă privind o relaţie curbilinie sau asupra influenţei avute de rezultatele atipice".

8 Regresia
Predicţia cu precizie

Rezumat

- Acolo unde există o relaţie între două variabile, sunt posibile estimarea sau predicţia scorului unei persoane la o variabilă pornind de la scorul obţinut la cealaltă variabilă. Cu cât este mai puternică corelaţia, cu atât mai bună este predicţia.
- Regresia poate fi folosită pe date asemănătoare celor utilizate pentru coeficienţii de corelaţie. Totuşi, este mult mai rar utilizată, parţial din cauza problemei de comparaţie între valorile obţinute pentru seturi diferite de variabile. (Mărimea lui beta poate fi folosită dacă este necesară o asemenea comparaţie.)
- Variabila dependentă într-o regresie este cea a cărei valoare va fi estimată. Este cunoscută, de asemenea, ca variabilă-criteriu, variabila prezisă sau variabila *Y*.
- Variabila independentă este variabila folosită pentru a realiza predicţia. Este cunoscută, de asemenea, ca variabila-predictor sau variabila *X*.
- Este nevoie de o mare atenţie pentru a nu confunda variabila independentă cu cea dependentă. Acest lucru se poate întâmpla uşor în cazul unei regresii simple. Cel mai bun mod de a evita aceste probleme este de a examina scatterplot-ul sau diagrama scatter a relaţiei dintre cele două variabile. Asiguraţi-vă că axa orizontală *X* este variabila independentă şi că axa verticală *Y* este variabila dependentă. Se poate investiga care este punctul de tăiere, aproximativ, de pe diagrama scatter şi să vă faceţi o idee asupra valorii pe care ar trebui să o aibă panta. Punctul de tăiere este punctul în care panta se intersectează cu axa verticală. Aceste estimări pot fi comparate cu valorile calculate pentru a ne asigura că nu există nici o eroare. Dacă apar probleme, motivul cel mai probabil este acela că variabilele dependente şi independente au fost încurcate.
- Tehnica simplă de regresie descrisă în acest capitol exprimă relaţiile la nivelul unităţilor de măsură originale ale variabilelor implicate. Deci, dacă două studii diferite folosesc variabile uşor diferite, este dificil să compari rezultatele folosind această formă de regresie.
- Într-o regresie, relaţia dintre două variabile este descrisă matematic de panta dreptei care se potriveşte punctelor diagramei de împrăştiere împreună cu punctul în care linia de regresie intersectează axa (verticală) a diagramei scatter. Prin urmare, relaţia dintre cele două variabile necesită valoarea pantei (în mod normal, simbolurile *B* sau *b*) şi aflarea punctului de tăiere la intersecţia cu axa verticală (în mod normal este dat simbolul *a* sau este descris ca o constantă).

- Regresia devine o tehnică mult mai importantă atunci când sunt folosite mai multe variabile pentru predicția valorilor unei alte variabile. Aceste tehnici sunt cunoscute sub numele de regresii multiple (vezi capitolele 28 și 29). Când variabila dependentă este o categorie nominală, atunci statistica potrivită va fi o formă de regresie logistică (vezi capitolul 32 și 33).

Vom ilustra procesarea unei regresii simple și a unei diagrame de regresie cu datele din tabelul 8.1, care prezintă scoruri ale abilităților muzicale și matematice pentru 10 copii. Aceste date sunt identice cu cele folosite în capitolul anterior în cazul corelației. În acest fel, veți putea aprecia mai ușor diferențele dintre corelație și regresie.

Tabelul 8.1. *Scorurile abilităților muzicale și matematice pentru 10 copii*

Music score	Mathematics score
2	8
6	3
4	9
5	7
7	2
7	3
2	9
3	8
5	6
4	7

Scorurile abilităților muzicale formează criteriul sau variabila dependentă, iar cele ale abilităților matematice formează predictorul sau variabila independentă. În cazul regresiei, este esențial să asociem criteriului sau variabilei dependente axa verticală (axa Y) a scatterplot-ului, și predictorului sau variabilei independente axa orizontală (axa X).

8.1. Introducerea datelor

Dacă ați salvat datele, selectați fișierul respectiv. Dacă nu, introduceți datele încă o dată.

Pasul 1

În „Variable View" din „Data Editor" denumiți primul rând „Music" și al doilea „Math".

Înlăturați cele două zecimale schimbând cifra prezentă aici în zero.

Name	Type	Width	Decimals
Music	Numeric	8	0
Maths	Numeric	8	0

Pasul 2

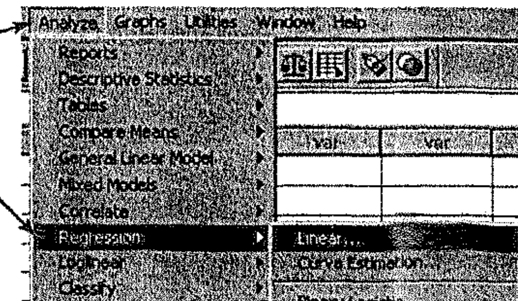
În „Data View" din „Data Editor" introduceți datele pentru muzică în prima coloană și cele pentru matematică în cea de-a doua coloană.

	Music	Maths
1	2	8
2	6	3
3	4	9
4	5	7
5	7	2
6	7	3
7	2	9
8	3	8
9	5	6
10	4	7

8.2. Regresia simplă

Pasul 1

Selectați „Analyze", „Regression" și „Linear...".

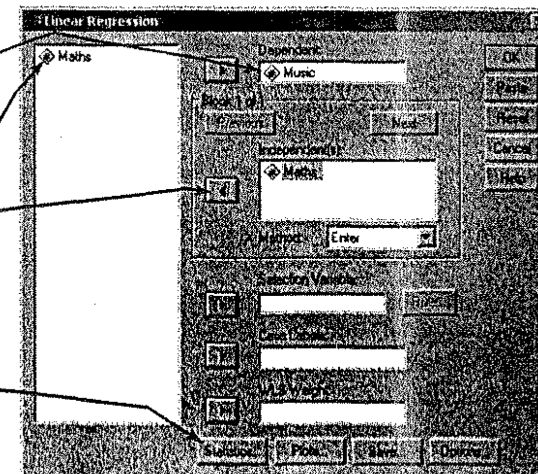


Pasul 2

Selectați „Music" și apoi apăsați butonul ▶ de lângă eticheta „Dependent:" pentru a introduce această variabilă în căsuța variabilei dependente.

Selectați „Math" și apoi butonul ◀ de lângă eticheta „Independent(s):" pentru a introduce această variabilă în căsuța variabilei(lor) independente.

Selectați „Statistics...".

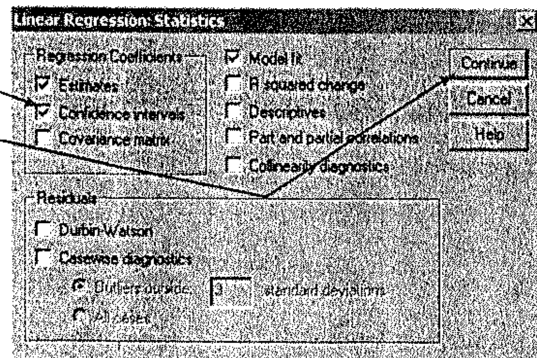


Pasul 3

Selectați „Confidence Intervals".

Selectați „Continue".

Apăsați „OK" din ecranul precedent, care reapare.



8.3. Interpretarea output-ului

Acest ultim tabel cu rezultate conține detaliile esențiale pentru analiza regresiei. Este foarte ușor de inversat accidental variabila independentă cu cea dependentă. Verificați tabelul intitulat „Coefficients[a]". Sub acest tabel este afișat numele variabilei dependente. În acest caz, „Music" este variabila dependentă. Dacă a fost citit „Math" atunci ați făcut o greșeală și analiza trebuie refăcută pentru că valorile regresiei ar fi incorecte.

Constanta este 8,425. Acesta este punctul în care linia de regresie intersectează axa verticală.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	8.425	.725		11.620	.000	6.753	10.097
	Maths	-.633	.109	-.900	-5.832	.000	-.883	-.383

a. Dependent Variable: Music

Coeficientul nestandardizat al regresiei este –0,633. Acesta semnifică faptul că, pentru fiecare creștere cu 1 a variabilei „Math", valoarea variabilei „Music" descrește cu –0,633.

Coeficientul standardizat al regresiei este –0,900. Acesta este mai mult sau mai puțin coeficientul de corelație Pearson dintre abilitățile muzicale și cele matematice.

Intervalul de încredere ia valori de la –0,833 la –0,383. Însemnând că este foarte probabil ca coeficientul nestandardizat al populației să aibă valori cuprinse între –0,833 și –0,383.

În regresii simple care implică două variabile, în mod convențional se reportează ecuația de regresie ca o pantă (b) și ca un punct de intersecție (a) (vezi *ISP*, capitolul 8). SPSS-ul nu

urmează această terminologie. Din păcate, la acest moment output-ul aplicație SPSS este mult mai complex și mai detaliat decât sofisticarea statistică observată la majoritatea studenților:

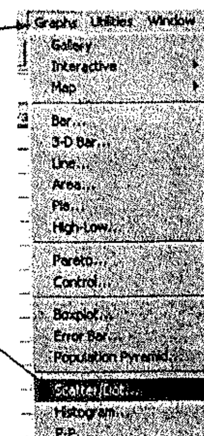
- B este panta. Panta liniei de regresie este denumită coeficient de regresie nestandardizat în SPSS. Coeficientul de regresie nestandardizat dintre variabilele „Music" și „Math" este afișat sub B și are valoarea –0,633, ceea ce, rotunjit la două zecimale, este –0,63. Acesta semnifică faptul că pentru fiecare creștere cu 1,00, a axei orizontale valoarea de pe axa verticală se modifică cu –0,633.
- Intervalul de încredere de 95% pentru acești coeficienți este de la –0,88 (–0,883) la –0,38 (–0,383). Deoarece regresia este bazată pe un eșantion, și nu pe întreaga populație, există întotdeauna un risc ca acest coeficient de regresie obținut pe eșantion să difere de cel al populației. Intervalul de încredere de 95% arată intervalul pantelor de regresie în care putem fi siguri într-o proporție de 95% că panta pentru populație se va găsi.
- Punctul de intersecție a este identificat în SPSS ca o constantă. Intersecția este prezentată ca fiind „Constant" și este 8,425, ceea ce, rotunjit la două zecimale, este 8,43. Este punctul în care linia de regresie intersectează axa verticală y.
- Intervalul de încredere pentru intersecție este de la 6,753 la 10,097. Ceea ce înseamnă că, în funcție de eșantion, este 95% probabil ca intersecția populației să se afle în intervalul de la 6,75 la 10,10.
- Coloana numită „Beta" are valoarea –0,900. Aceasta este de fapt corelația Pearson între cele două variabile. Cu alte cuvinte, dacă transformați valorile avute în valori standard (scoruri z), panta regresiei și coeficientul de corelație sunt același lucru.

8.4. Scatterplot-ul de regresie

Este în general recomandabilă cercetarea diagramei scatter pentru cele două variabile atunci când se realizează o regresie. Urmați pașii implicați în redactarea unei diagrame scatter așa cum sunt descriși în capitolul 7.

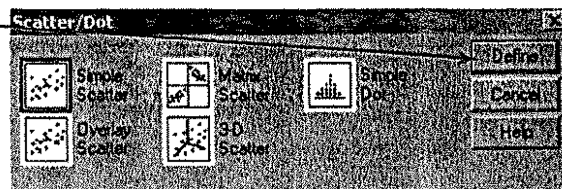
Pasul 1

Selectați „Graphs" și „Scatter/Dot...".



Pasul 2

Selectați „Define" din moment ce „Simple" este deja selectat. Se numește opțiunea predefinită.



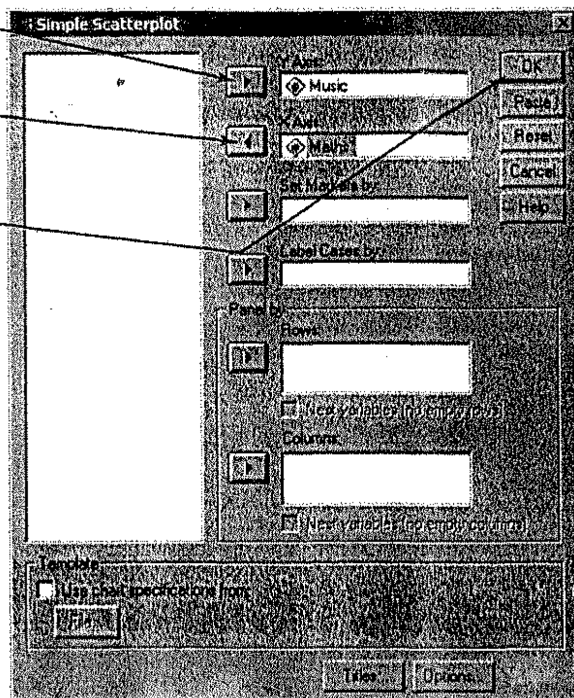
Pasul 3

Selectați „Music" și apoi apăsați butonul ▶ de lângă eticheta „Y Axis:" pentru a introduce această variabilă în căsuța variabilei-criteriu.

Selectați „Math" și apoi apăsați butonul ◀ de lângă eticheta „X Axis:" pentru a introduce această variabilă în căsuța variabilei-predictor.

Apăsați „OK".

Output-ul va fi identic cu cel pentru corelație din secțiunea 7.9.

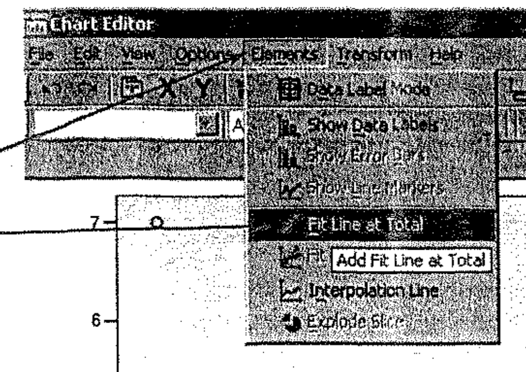


Pasul 4

Pentru a încadra o linie de regresie în scatterplot, dați dublu clic oriunde în interior și se va deschide „Chart Editor".

Selectați „Elements" și „Fit Line at Total".

Selectați „Close" din fereasta „Properties".

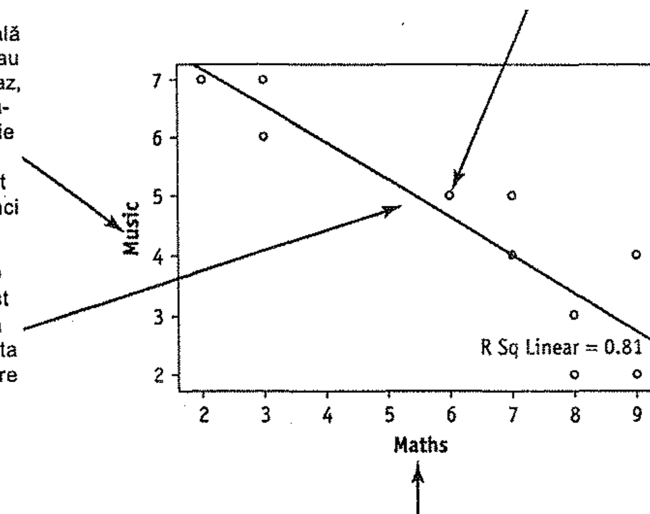


8.5. Interpretarea output-ului

Punctele de pe scatterplot sunt apropiate de linia de regresie (linia se potrivește bine cu datele). În plus, punctele par să formeze o linie dreaptă (relația nu este curbilinie). De asemenea, nu este nici un semn care să indice valori atipice – în special ridicate sau scăzute.

În regresie, axa verticală este variabila-criteriu sau dependentă. În acest caz, este „Music". Amestecarea axelor într-o regresie este o eroare des întâlnită. Dacă ați făcut această greșeală, atunci începeți încă o dată.

Linia de regresie are o pantă negativă în acest caz; de exemplu, de la stânga sus spre dreapta jos. În consecință, *B* are o valoare negativă.



În regresie, axa orizontală este variabila predictor sau independentă. În acest caz este „Math". Amestecarea axelor într-o regresie este o eroare des întâlnită. Dacă ați făcut această greșeală, atunci refaceți analiza.

- Linia de regresie oblică, de la stânga sus spre dreapta jos, indică o relaţie negativă între cele două variabile. Amintiţi-vă că numai în cazul în care axele intersectează în punctul zero atât axa verticală, cât şi pe cea orizontală, atunci interpretarea riscă să fie greşită.
- Punctele par relativ apropiate de linie, ceea ce sugerează că Beta (corelaţia) ar trebui să fie un număr mare (negativ) şi că intervalul de încredere este relativ mic.

8.6. Raportarea output-ului

Deşi tot output-ul din SPSS este pertinent pentru un utilizator sofisticat, mulţi dintre utilizatori ar prefera varianta simplificată la această etapă.

- Având acestea în minte, am putea scrie despre analiza din acest capitol în felul următor: „Diagrama scatterplot a relaţiei dintre abilităţile matematice şi muzicale sugerează o relaţie liniară negativă între cele două variabile. Este posibilă predicţia cu acurateţe a abilităţii muzicale a unei persoane cunoscând abilitatea sa matematică. Ecuaţia este $Y' = 8{,}43\ (+\ /\ -0{,}633X)$, unde X este valoarea individului pentru abilitatea matematică şi Y' cea mai bună predicţie a valorii pentru abilitatea muzicală."
- O alternativă ar fi afişarea scatterplot-ului şi scrierea dedesubt $a = 8{,}43$ şi $B = -0{,}63$.
- Pot fi adăugate şi intervalele de încredere, după cum urmează: „Intervalul de încredere 95% pentru panta liniei de regresie este de la -0,88 la -0,38. Deoarece acest interval de încredere nu include valoarea 0, panta diferă semnificativ faţă de o linie orizontală dreaptă". Totuşi, aceasta este o interpretare relativ sofisticată pentru un începător în domeniul statisticii.

9 Eşantioane şi populaţii
Generarea unui eşantion aleatoriu

Rezumat

- Eşantionarea aleatorie este un aspect-cheie al statisticii. Acest capitol explică cum poate fi generată rapid eşantionarea aleatorie. Aceasta nu este o procedură folosită des în analiza statistică, dar poate oferi o experienţă folositoare despre natura proceselor de randomizare.
- Se poate înţelege mai bine statistica inferenţială şi eşantionarea prin folosirea de eşantioane aleatorii de date pentru cercetarea variabilităţii rezultatelor pentru analize statistice viitoare cu aceste eşantioane aleatorii. Variaţia caracteristicilor eşantioanelor este cunoscută ca eroare de eşantionare şi este baza statisticii inferenţiale.
- Eşantionarea aleatorie poate fi utilizată cu seturi imense de date pentru a întreprinde analize preliminare. În trecut, când timpul de calculator era scump, iar viteza de procesare lentă, acestea erau de mare ajutor.
- Fiţi atenţi la căsuţele de dialog pe măsură ce parcurgeţi acest capitol. Veţi vedea opţiuni care vă permit selectarea eşantioanelor pe bază de alte criterii, cum ar fi datele la care participanţii au fost intervievaţi.

În acest capitol este ilustrată selecţia de eşantioane aleatorii dintr-un set de valori. Scopul de bază este acela de a permite celor care studiază statistica pentru prima dată să încerce eşantionarea aleatorie pentru a putea avea o imagine despre distribuţiile eşantionelor. Aceasta ar trebui să conducă spre o mai bună apreciere a estimării în statistică şi a fragilităţii care poate submina considerabil proceduri matematice greoaie. Vom ilustra generarea unui eşantion aleatoriu dintr-un set de date constând în scorurile la testul de extroversiune ale unui număr de 50 de piloţi de linie aeriană, prezentate în tabelul 4.1.

Tabelul 4.1. *Rezultatele obţinute la evaluarea extroversiunii pentru 50 de piloţi de linie aeriană*

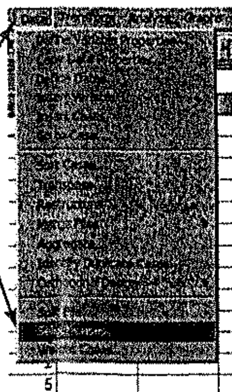
3	5	5	4	4	5	5	3	5	2
1	2	5	3	2	1	2	3	3	3
4	2	5	5	4	2	4	5	1	5
5	3	3	4	1	4	2	5	1	2
3	2	5	4	2	1	2	3	4	1

9.1. Selectarea unui eşantion aleatoriu

Pasul 1

Dacă aţi salvat datele, selectaţi fişierul. Dacă nu, introduceţi datele din nou.

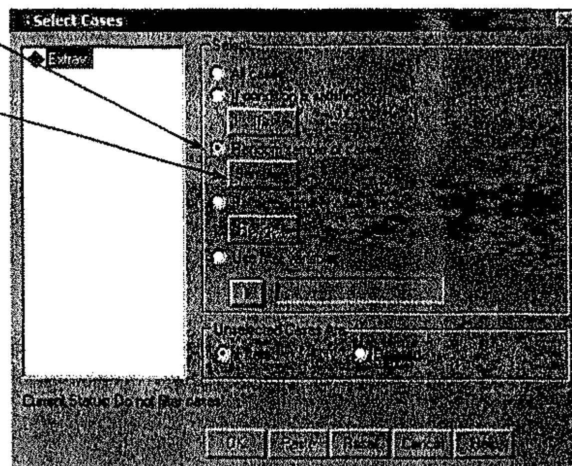
Selectaţi „Data" şi „Select Cases...".



Pasul 2

Selectaţi „Random sample of cases".

Selectaţi „Sample".



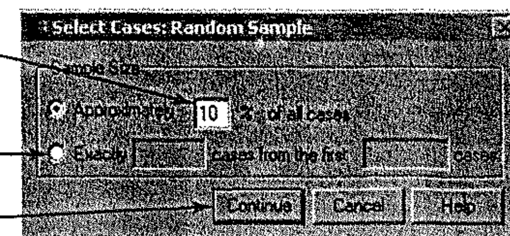
Pasul 3

Pentru a selecta un procent de aproximativ 10% din toate cazurile, introduceţi 10 în căsuţa de lângă „Approximately".

Pentru a selecta exact cinci cazuri, selectaţi „Exactly", introduceţi „5" cazuri din primele „50".

Selectaţi „Continue".

Apăsaţi „OK" din ecranul precedent, care reapare.



9.2. Interpretarea rezultatelor

Cazurile care au fost selectate nu au o linie diagonală peste numărul cazului, şi au cifra „1" în noua coloană creată, denumită „filter_$".

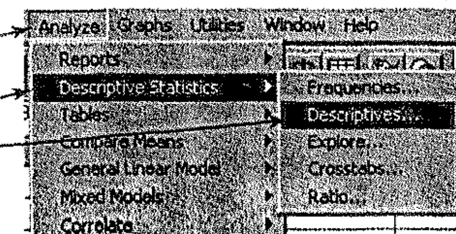
Cazurile care nu au fost selectate au o linie diagonală peste numărul cazului, şi cifra 0 în noua coloană creată, denumită „filter_$".

	Extrav	filter_$
1	3	0
2	5	0
3	5	0
4	4	0
5	4	0
6	5	0
7	5	0
8	3	0
9	5	1
10	2	0
11	1	0
12	2	1
13	5	0
14	3	0

9.3. Analiza statistică pe un eşantion aleatoriu

Pasul 1

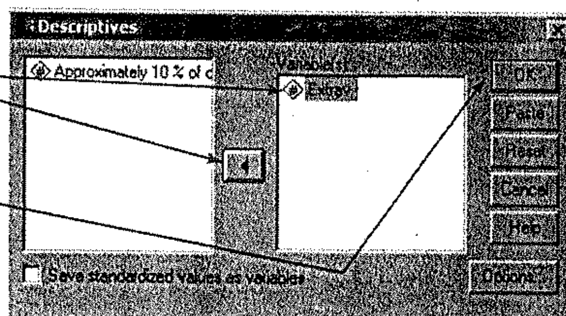
Pentru a analiza acest eşantion, selectaţi „Analyze" şi apoi analiza dorită cum ar fi „Descriptive Statistics" şi „Descriptives...".



Pasul 2

Selectați „Extrav" și apăsați butonul ◄ pentru a introduce variabila „Extrav" în lista de variabile.

Apăsați „OK".



Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Extrav	7	1	5	3.29	1.496
Valid N (listwise)	7				

Observați că au fost selectate 7 cazuri. Dacă s-ar fi cerut exact 10%, atunci 5 cazuri din 50 ar fi trebuit specificate la pasul 3 de la secțiunea 9.1.

10 Selectarea cazurilor

Rezumat

- Acest capitol explică maniera de selectare a unui anumit subgrup din eșantion. De exemplu, veți dori să analizați datele referitoare numai la tineri sau numai la femei.
- Este posibil să selectați subgrupuri bazându-vă pe criterii multiple.
- Câteodată, utilizarea subgrupurilor conduce la o mult mai clară înțelegere a tendințelor existente în date, decât ar fi fost posibilă dacă, de exemplu, s-ar fi folosit tabelele de asociere după sex.

Câteodată vom dori să realizăm calcule pe subgrupuri din eșantionul nostru. De exemplu, am putea dori să corelăm abilitățile muzicale și cele matematice la fete și băieți separat (*a*), la copiii mari și la cei mici separat (*b*), la fetele și băieții mari și mici separat (*c*). Pentru a face acest lucru, trebuie să avem un cod pentru gen și vârstă, cum ar fi „1" pentru fete și „2" pentru băieți. De asemenea, trebuie să decidem care este vârsta de separare pentru a determina cum sunt încadrați copiii în grupul cu vârsta mai mică și în grupul cu vârsta mai mare. Vom folosi vârsta de 10 ani ca limită fixă, copiii de 9 ani sau mai puțin încadrându-se în grupul cu vârstă mică, iar cei de 10 ani sau mai mulți încadrându-se în grupul cu vârstă mare. Apoi va trebui să selectăm pe rând cele două grupuri și să efectuăm calculele. Vom ilustra această selecție a cazurilor folosind datele din tabelul 10.1, care prezintă rezultatele la muzică și matematică pentru zece copii, alături

Tabelul 10.1. *Valorile numerice pentru abilitățile la muzică și matematică pentru 10 copii, împreună cu codul pentru sex și vârstă*

Music score	Mathematics score	Sex	Age
2	8	1	10
6	3	1	9
4	9	2	12
5	7	1	8
7	2	2	11
7	3	2	13
2	9	2	7
3	8	1	10
5	6	2	9
4	7	1	11

de codul pentru gen şi vârstă (valorile pentru abilităţile muzicale şi matematice sunt identice cu cele prezentate anterior în tabelul 7.1).

Evident, selecţia limitei fixe este importantă. Trebuie să fiţi atenţi la excluderea neprevăzută a unor cazuri.

10.1. Introducerea datelor

Pasul 1

Dacă aţi salvat fişierul de date de la capitolul 7, selectaţi acel fişier şi definiţi noi variabile în „Variable View" din „Data Editor".

Dacă nu, definiţi-vă cele patru variabile.

Înlăturaţi cele două zecimale.

Salvaţi acest fişier pentru utilizare în capitolele 15 şi 16.

	Name	Type	Width	Decimals
1	Music	Numeric	8	0
2	Maths	Numeric	8	0
3	Sex	Numeric	8	0
4	Age	Numeric	8	0

Pasul 2

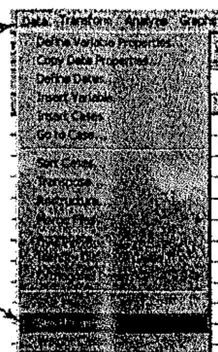
Introduceţi datele în „Data View" din „Data Editor".

	Music	Maths	Sex	Age
1	2	8	1	10
2	6	3	1	9
3	4	9	2	12
4	5	7	1	8
5	7	2	2	11
6	7	3	2	13
7	2	9	2	7
8	3	8	1	10
9	5	6	2	9
10	4	7	1	11

10.2. Selectarea cazurilor

Pasul 1

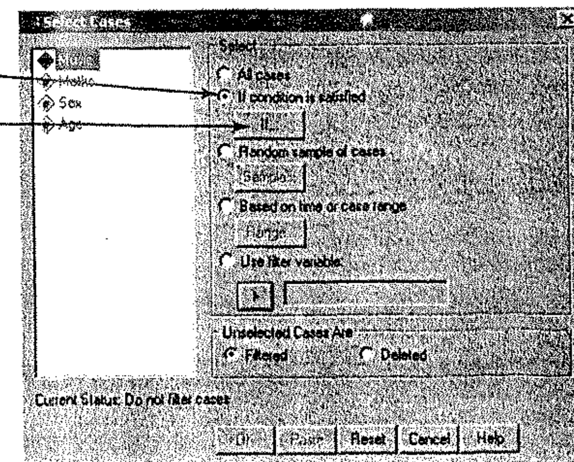
Selectaţi „Data" şi „Select Case...".



Pasul 2

Selectaţi „If condition is satisfied".

Selectaţi „If".



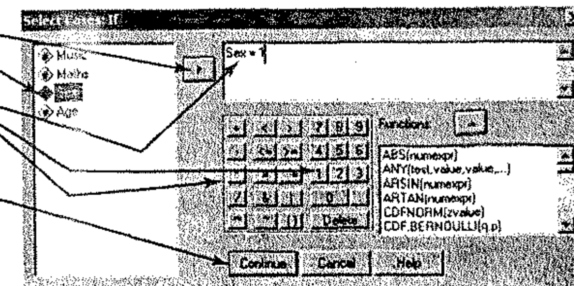
Pentru a selecta fetele, selectaţi „Sex" şi apoi butonul ▶

Selectaţi „=" şi „1". Alternativ introduceţi „Sex = 1".

Selectaţi „Continue".

Selectaţi „OK" din ecranul precedent, care reapare.

Duceţi la bun sfârşit analiza dorită (de exemplu, corelaţia).

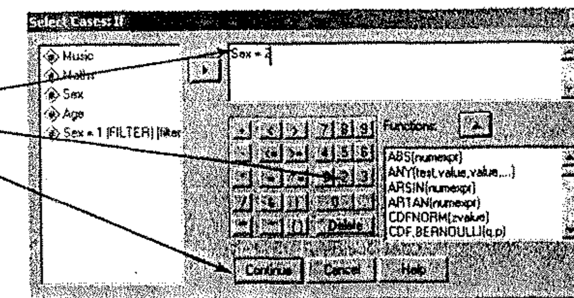


Pentru a selecta băieţii, selectaţi „Data", „Select Cases..." şi „If".

Evidenţiaţi „1" şi înlocuiţi-l cu „2".

Selectaţi „Continue" şi apoi „OK".

Faceţi analiza (de exemplu, corelaţia).



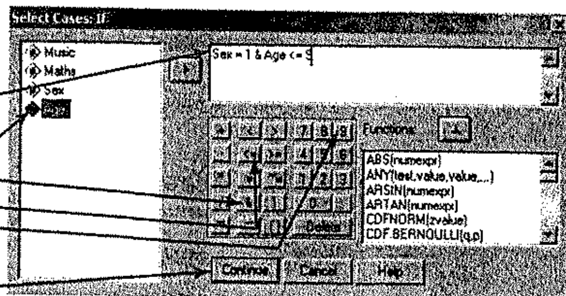
Pentru a selecta fetele de 9 sau mai puţini ani, selectaţi „Data", „Select Cases..." şi „If"

Evidenţiaţi „2" şi înlocuiţi-l cu „1".

Selectaţi „&", „Age", butonul ▶, „<=" (mai mic sau egal) şi „9". Sau introduceţi de la tastatură.

Selectaţi „Continue" şi apoi „OK".

Realizaţi analiza (de exemplu, corelaţia).



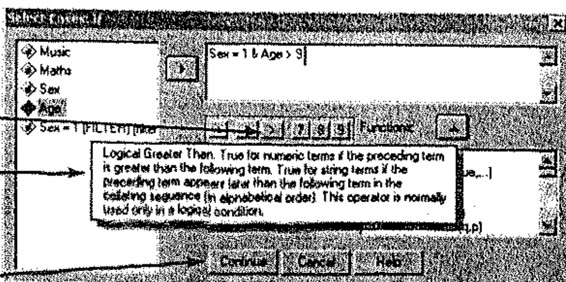
Pentru a selecta fetele de mai mult de 9 ani, selectaţi „Data", „Select Cases..." şi „If".

Marcaţi „<=" şi înlocuiţi-l cu „>" (semnul „mai mare").

Pentru a vizualiza sau a aminti semnificaţia semnului, deplasaţi cursorul pe el şi daţi clic dreapta.

Selectaţi „Continue" şi apoi „OK".

Realizaţi analiza (de exemplu, corelaţia).



11 Eroarea standard

Rezumat

- Eroarea standard este un indice pentru variabilitatea mediilor în cazul mai multor eşantioane extrase din populaţie. Nu este nimic altceva decât abaterea standard, dar aplicată eşantioanelor în loc de scoruri. Cu alte cuvinte, eroarea standard este media măsurii cu care media eşantioanelor diferă faţă de media populaţiei din care au fost extrase.
- Este folosită în cea mai mare parte ca un pas intermediar în alte tehnici statistice, cum ar fi Testul *t*. Nu este mare nevoie de calculul erorii standard de sine stătător, din moment ce aduce foarte puţin sau nimic nou faţă de calculul abaterii standard sau al varianţei.
- Cu toate acestea, poate fi folosită la fel ca abaterea standard şi varianţa ca un indice pentru variabilitatea scorurilor unei variabile. Un cercetător cu experienţă găseşte varianţa, abaterea standard şi eroarea standard la fel de folositoare ca indicatori ai variabilităţii scorurilor.
- Eroarea standard este un concept important pentru calculul intervalelor de încredere. Într-adevăr, apare în multe contexte în care atenţia este îndreptată spre caracteristicile eşantioanelor mai degrabă decât spre cele ale scorurilor.
- Există două versiuni ale erorii standard. Eroarea standard aplicată unui set de scoruri şi cea estimată, care este folosită atunci când se încearcă estimarea erorii standard a populaţiei din cea a unui eşantion. De fapt, SPSS calculează numai eroarea standard estimată. De aici încolo, pentru eroarea standard în ouput-ul SPSS-ului citiţi eroarea standard estimată.

Calcularea erorii standard estimate pentru medie este ilustrată cu un set de şase scoruri ale respectului de sine, prezentate în tabelul 11.1.

Tabelul 11.1. *Date pentru exemplul de eroare standard*

Self-esteem [illegible]	
5	6
[illegible]	[illegible]
[illegible]	5

11.1. Introducerea datelor

Introduceți datele în „Data Editor"

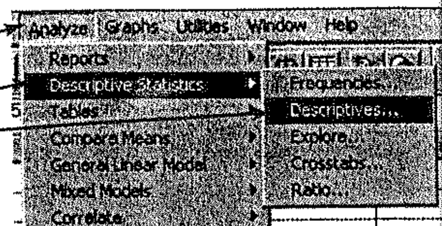
Etichetați variabila „Esteem".

	Esteem
1	5
2	7
3	3
4	6
5	4
6	5
7	

11.2. Eroarea standard estimată a mediei

Pasul 1

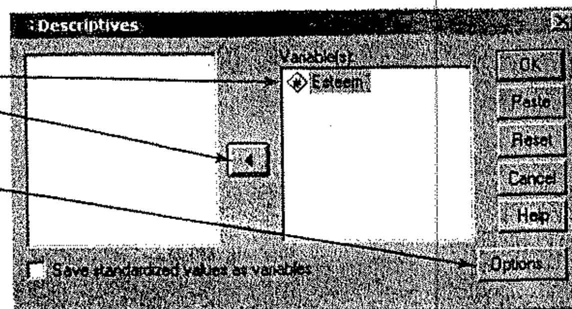
Selectați „Analyze", „Descriptive Statistics" și „Descriptive...".



Pasul 2

Selectați „Esteem" și apoi butonul ◄ pentru a introduce variabila „Esteem" în „Variabile(s):".

Selectați „Options...".

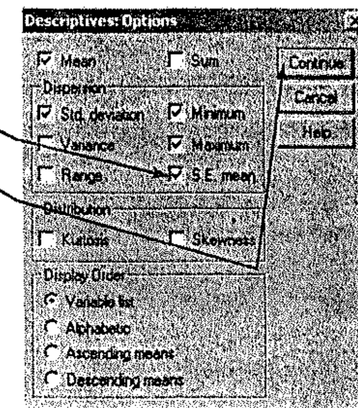


Pasul 3

Selectați „S.E. mean".

Selectați „Continue".

Apăsați „OK" din ecranul anterior, care reapare.



11.3. Interpretarea output-ului

Descriptive Statistics

	N	Minimum	Maximum	Mean		Std.
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic
Esteem	6	3	7	5.00	.577	1.414
Valid N (listwise)	6					

Eroarea standard (estimată) pentru medie a acestui eșantion este 0,577. Este o indicație a mediei cu care mediile eșantioanelor diferă de media populației din care au fost extrase.

- Tabelul oferă valoarea erorii standard a mediilor eșantioanelor ca fiind 0,58, fiind rotunjită la două zecimale. Aceasta este valoarea medie cu care mediile eșantioanelor ($N = 6$) diferă față de media populației.
- Este o estimare bazată pe un eșantion și ar trebui denumită estimarea erorii standard.
- Tabelul include și alte informații cum ar fi media (5,00), abaterea standard estimată a populației bazată pe acest eșantion, valorile minime și maxime ale datelor.
- Ultima coloană oferă abaterea standard (estimată) a celor șase scopuri, care este 1,41.

11.4. Raportarea output-ului

În general, în statistica psihologică nu se raportează eroarea standard pentru mediile eșantioanelor ca un fapt în sine. Ar fi mai obișnuită raportarea ca parte a anumitor teste de semnificație. Totuși, în multe circumstanțe este la fel de informativă ca și varianța sau abaterea standard a unui eșantion, pentru că poartă o relație simplă cu ambele.

12 Testul *t*
Compararea a două eşantioane de scoruri corelate/relaţionate

Rezumat

- Testul *t* este folosit pentru evaluarea semnificaţiei statistice a diferenţei dintre mediile pentru două seturi de scoruri. Astfel, ajută la elucidarea întrebărilor comune ale cercetătorilor dacă valoarea medie pentru un set de scoruri diferă de valoarea medie pentru alt set de scoruri.
- Semnificaţia statistică este evaluată folosind variabilitatea în datele disponibile pentru a evalua probabilitatea ca cele două medii să fie diferite dacă nu ar exista nici o deosebire fundamentală între cele două eşantioane. Ipoteza de nul sugerează că nu este nici o diferenţă între mediile eşantioanelor. Despre diferenţele între mediile eşantioanelor, care ar fi foarte neobişnuite, aleatorii, dacă ipoteza de nul ar fi adevărată, se spune că sunt semnificative statistic. Ele ne ghidează spre acceptarea ipotezei conform căreia există o diferenţă reală între cele două medii şi este puţin probabil ca aceasta să fie un efect aleatoriu datorat eşantionării.
- Pentru că în cercetare, în mod invariabil, se lucrează cu eşantioane de oameni extrase din populaţia latentă, trebuie să estimăm dacă orice diferenţă pe care o obţinem între cele două seturi de valori este semnificativă statistic. Cu alte cuvinte, este diferenţa de medie obţinută atât de diferită de o diferenţă zero încât să fie puţin probabil ca eşantioanele să provină din aceeaşi populaţie?
- Există două variante al Testului *t*. Una este folosită când cele două seturi de scoruri ce trebuie comparate provin dintr-un singur eşantion sau când coeficientul de corelaţie între cele două seturi este mare. Această variantă este cunoscută ca Testul *t* pentru eşantioane perechi. Cealaltă versiune a Testului *t* este utilizată în momentul în care două seturi diferite de valori provin din grupuri diferite de participanţi. Faceţi referinţă la capitolul următor dacă cele două medii pe care doriţi să le comparaţi provin din grupuri distincte de participanţi.
- Dacă aţi folosit o procedură de echivalare pentru a face perechi de oameni cu caracteristici similare, atunci veţi folosi, de asemenea, Testul *t* pentru eşantioane perechi din acest capitol – în mod special în cazul în care cele două seturi de valori sunt corelate semnificativ.
- Introducerea datelor pentru variabilele perechi şi independente (neperechi) este foarte diferită în SPSS, deci aveţi grijă să planificaţi analiza înainte de introducerea datelor pentru a evita problemele şi munca inutilă.
- Dacă aveţi mai mult de două seturi de valori de comparat, faceţi referire la capitolul 21 la analiza de varianţă cu eşantioane perechi.

- Testul *t* descris în acest capitol este cunoscut ca Testul *t* pentru eşantioane perechi. Fundamental, aceasta înseamnă că cele două seturi de scoruri provin dintr-un singur eşantion de participanţi. Ar trebui să existe o corelaţie între valorile celor două măsurări deoarece, din alt punct de vedere, aceste două seturi de scoruri nu sunt corelate.
- Testul *t* pentru eşantioane perechi lucrează optim dacă distribuţia diferenţelor dintre cele două seturi de valori se prezintă, aproximativ, sub formă de clopot (asta se întâmplă dacă avem de-a face cu o distribuţie normală). Dacă distribuţia este foarte diferită de forma de clopot, ar trebui luată în considerare utilizarea unei tehnici statistice relaţionate nonparametrică pentru eşantioane perechi, cum ar fi testul de perechi Wilcoxon (vezi capitolul 18).

Procesarea unui Test *t* este ilustrată cu datele din tabelul 12.1, care arată numărul de contacte vizuale avute de aceiaşi bebeluşi cu mamele lor la 6 şi 9 luni. Scopul acestei analize este de a vedea dacă numărul contactelor vizuale se modifică între aceste vârste.

Tabelul 12.1. *Numărul segmentelor de un minut cu contact vizual la vârste diferite*

Baby	6 months	9 months
Clara	3	7
Martin	5	6
Sally	5	3
Angie	4	8
Trevor	3	5
Sam	7	9
Bobby	8	7
Sid	7	9

12.1. Introducerea datelor

Pasul 1

În „Variable View" din „Data Editor" etichetaţi primul rând „Six_mths" şi al doilea rând „Nine_mths".

Înlăturaţi cele două zecimale, schimbând cifra prezentă cu 0.

	Name	Type	Width	Decimals
1	Six_mths	Numeric	8	0
2	Nine_mths	Numeric	8	0

Pasul 2

În „Data View" din „Data Editor" introduceţi datele în primele două coloane.

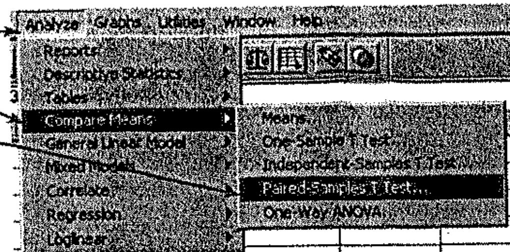
(Salvaţi acest fişier pentru utilizare în capitolul 18.)

	Six_mths	Nine_mths
1	3	7
2	5	6
3	5	3
4	4	8
5	3	5
6	7	9
7	8	7
8	7	9

12.2. Testul *t* pentru eşantioane perechi

Pasul 1

Selectaţi „Analyze", „Compare Means" şi „Paired-Samples T Test...".



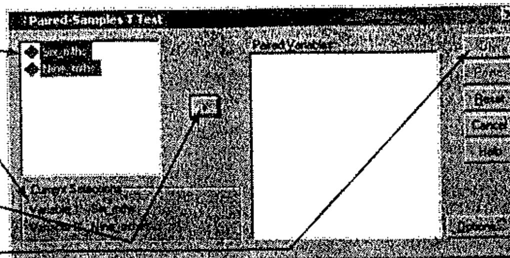
Pasul 2

Selectaţi „Six_mths" şi puneţi această variabilă lângă eticheta „Variable 1:" sub „Current Selections".

Selectaţi „Nine_mths" şi puneţi această variabilă lângă eticheta „Variable 2:" sub „Current Selections".

Apăsaţi butonul ▸ pentru a pune aceste două variabile în lista variabilelor relaţionate.

Apăsaţi „OK".



12.3. Interpretarea output-ului

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Six_mths	5.25	8	1.909	.675
	Nine_mths	6.75	8	2.053	.726

Primul tabel arată media, numărul de cazuri şi abaterea standard pentru cele două grupuri. Media pentru „Six_mths" este 5,25 şi abaterea standard este1,909. Cele două abateri standard sunt foarte similare, ceea ce este un avantaj.

Paired Samples Correlations

	N	Correlation	Sig.
Pair 1 Six_mths & Nine_mths	8	.419	.301

Al doilea tabel arată gradul în care cele două seturi de valori sunt corelate. Corelaţia dintre ele este 0,419. Aceasta este o corelaţie medie, şi totuşi nesemnificativă, pentru că nivelul de semnificaţie este mai mare ca 0,05. Testele relaţionate cum ar fi Testul *t* pentru eşantioane perechi ar trebui să aibă o corelaţie substanţială între cele două seturi de scoruri. Fiţi atenţi, acest tabel nu este testul de semnificaţie. Este o greşeală comună, printre începători, să se confunde această corelaţie dintre cele două variabile cu semnificaţia diferenţei dintre cele două variabile.

Primele trei coloane ce conţin cifre sunt componentele fundamentale ale calculului unui Test *t* relaţionat. Media de –1,500 este de fapt diferenţa dintre mediile pe 6 şi 9 luni, deci în realitate este media diferenţei. Valoarea lui *t* este bazată pe această medie a diferenţei (–1,500), divizată cu eroarea standard a mediei (0,756). Calculul oferă valoarea lui *t* (–1,984).

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference: Lower	Upper			
Pair 1 Six_mths – Nine_mths	–1.500	2.138	.756	–3.287	.287	–1.984	7	.088

Al treilea şi ultimul tabel oferă valoarea pentru *t* (–1,984), gradele de libertate (7) şi nivelul de semnificaţie, two-tailed (0,088). Pentru că nivelul de semnificaţie este mai mare ca 0,05, această diferenţă nu este semnificativă. Nivelul one-tailed este obţinut prin divizare cu 2, şi este 0,044, fiind semnificativ. Totuşi, dacă diferenţa nu este evaluată anterior colectării datelor pe baza unor raţiuni teoretice şi/sau empirice solide, numai Testul two-tailed este potrivit.

- În primul tabel din output, media numărului de contacte vizuale la 6 luni („Six_mths") și la 9 luni („Nine_mths") este afișată sub „Mean. Așadar, media numărului de contacte vizuale este 5,25 la 6 luni și 6,75 la 9 luni.
- Al doilea tabel de la output conține coeficienții de corelație (Pearson) între cele două variabile (de contacte vizuale la 6 luni și de contacte vizuale la 9 luni). În mod ideal, această valoare ar trebui să fie considerabilă (în realitate, este 0,419) și semnificativă sta-tistic (ceea ce nu este de fapt, având un nivel de semnificație two-tailed de 0,301). Testul *t* pentru eșantioane perechi presupune ca cele două variabile să fie corelate, și ați putea descoperi că Testul *t* pentru eșentioane independente (capitulul 13) este mai potrivit pentru acest caz.
- În al treilea tabel de la output, diferențele dintre cele două valori medii sunt prezentate pe coloana „Mean" de la „Paired Differences" și eroarea standard a acestei medii pe coloana „Std. Error Mean". Diferența dintre cele două medii este -1,50 și eroarea standard estimată mediilor pentru acest eșantion este 0,76.
- Valoarea *t* a diferenței dintre mediile celor două eșantioane, gradele sale de libertate și nivelul de semnificație two-tailed sunt, de asemenea, în acest tabel. Valoarea *t* este -1,984, care are un nivel de semnificație (two-tailed) de 0,088 cu șapte grade de libertate.

12.4. Raportarea output-ului

- Am putea prezenta aceste rezultate după cum urmează: „Media numărului de contacte vizuale la 6 luni ($M = 5{,}25$, $SD = 1{,}91$) și la 9 luni ($M = 6{,}75$, $SD = 2{,}05$) nu diferă semnificativ ($t = -1{,}98$, $DF = 7$, two-tailed $p = 0{,}088$)".
- Pentru a fi consecvenți, în această carte vom prezenta nivelul de probabilitate exact, pentru rezultate nesemnificative precum cele de mai sus. Totuși, este în egală măsură acceptată raportarea ca „$p > 0{,}05$" sau *NS* (prescurtare de la nesemnificativ).
- Observați că rezultatele ar fi fost statistic semnificative pentru un test de o singură ipoteză (one-tailed). Totuși, acest lucru ar fi trebuit să fie anticipat cu anumite argumente solide înainte de aflarea datelor. În acest caz s-ar fi putut scrie despre efect: „Cele două medii diferă semnificativ într-o direcție anticipată ($t = 1{,}98$, $DF = 7$, p (one-tailed) $= 0{,}044$)".
- Încă o dată, pentru a fi consecvenți în această carte vom raporta nivelul exact de semnificație pentru rezultatele semnificative, acolo unde este posibil. Observați că SPSS-ul afișează nivelul de semnificație ca „,000", deci va trebui să-l prezentăm ca „$p < 0{,}001$", din moment ce nivelul exact nu este oferit. Este în egală măsură acceptată raportarea probabilităților semnificative ca „$p < 0{,}005$", „$p < 0{,}01$" și „$p < 0{,}001$", după caz.
- Dacă preferați să utilizați intervalele de încredere, va trebui să prezentați rezultatele ca: „Media numărului de contacte vizuale la 6 luni este 5,25 (SD = 1,91) și la 9 luni 6,75 (SD = 2,05)". Diferența este 1,50. Intervalul de încredere 95% pentru această diferență este de la -3,29 la 0,29. Deoarece intervalul de încredere trece prin 0,00, diferența nu este statistic semnificativă la nivelul de semnificație two-tailed de 5%.
- Unii statisticieni susțin mai mult raportarea cu intervale de încredere decât cea cu nivele de semnificație. Totuși, rămâne o metodă neobișnuită prezentarea intervalelor de încredere.

13 Testul *t*
Compararea a două eșantioane de scoruri necorelate/nerelaționate

Rezumat

- Testul *t* pentru eșantioane independente este utilizat pentru a calcula dacă mediile pentru două seturi de variabile sunt diferite semnificativ una față de cealaltă. Este versiunea de Test *t* cel mai des folosită. Semnificația statistică presupune că cele două eșantioane diferă până la un punct, această diferență nefiind datorată întâmplării și nefiind o consecință a eșantionării. Variabilitatea inerentă a acestor variabile de date este utilizată pentru a estima probabilitatea apariției oricărei diferențe între cele două medii, dacă nu ar fi nici o diferență între cele două eșantioane.
- Testul *t* pentru eșantioane independente este utilizat atunci când cele două seturi de variabile provin din două eșantioane diferite de oameni. (Revedeți la capitolul anterior pentru Testul *t* relațional dacă valorile provin dintr-un singur set de oameni sau dacă ați folosit o procedură de echivalare.)
- Introducerea datelor pentru variabilele relaționate și nerelaționate este foarte diferită în SPSS, deci aveți grijă să planificați analiza înaintea introducerii datelor pentru a evita problemele și munca inutilă. Totuși, SPSS-ul este foarte flexibil și erorile de acest tip sunt de obicei ușor de corectat folosind caracteristici de genul celor „cut-and-paste".
- Procedurile SPSS al Testului *t* pentru eșantioane independente sunt foarte folositoare și depășesc tratarea lor obișnuită de manual. Astfel, acestea includ o opțiune pentru calcularea Testului *t* atunci când varianța celor două eșantioane de valori diferă semnificativ una față de cealaltă. Majoritatea manualelor sugerează eronat faptul ca Testul *t* ar fi imprecis atunci când varianțele celor două grupuri sunt inegale. Această versiune adițională a Testului *t* pentru eșantioane independente este rareori menționată în manualele statistice, dar este extrem de utilă.
- Dacă aveți mai mult de două seturi de valori de comparat, faceți referire la capitolul 20 la analiza pentru eșantioane de variație independente.

Procesarea unui Test *t* pentru eșantioane independente este ilustrată cu datele din tabelul 13.1, care arată valorile la un test de emotivitate pentru 12 copii ce provin din familii cu doi părinți și 10 copii ce provin din familii cu un părinte. În SPSS acest tip de Test *t* este denumit Testul *t* pentru eșantioane independente. Scopul analizei este de a aprecia dacă valorile emotivității sunt diferite pentru familiile cu doi părinți față de cele cu un singur părinte.

Tabelul 13.1. *Scorurile emotivității pentru copiii proveniți din familii cu doi părinți și cu un singur părinte*

Two-parent family X_1	Lone-parent family X_2
12	6
18	9
14	4
10	13
19	14
8	9
15	8
11	12
10	11
13	9
15	
16	

13.1. Introducerea datelor

Pasul 1

În „Variable View" din „Data Editor" etichetați primul rând, „Family". Acesta va defini cele două tipuri de familii.

Etichetați al doilea rând, „Emotion". Acestea sunt rezultatele la testul de emotivitate.

	Name	Type	Width	Decimals
1	Family	Numeric	8	0
2	Emotion	Numeric	8	0

Înlăturați cele două zecimale, înlocuind cifra existentă cu zero.

Pasul 2

În „Data View" din „Data Editor" introduceți valorile pentru cele două variabile în primele două coloane.

Salvați acest fișier pentru utilizarea ulterioară în capitolul 18.

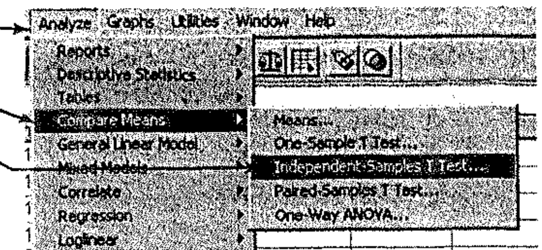
	family	emotion
1	2	12
2	2	18
3	2	14
4	2	10
5	2	19
6	2	8
7	2	15
8	2	11
9	2	10
10	2	13
11	2	15
12	2	16
13	1	6
14	1	9
15	1	4
16	1	13
17	1	14
18	1	9
19	1	8
20	1	12
21	1	11
22	1	9

Priviți cu atenție pasul 2. Observați că sunt două coloane de date. A doua coloană („Emotion") conține 22 de valori ale testului de emotivitate de la ambele grupe de copii. Datele nu sunt păstrate separat pentru cele două grupe; valorile „1" din prima coloană indică, în exemplul nostru, copiii proveniți din familii cu un singur părinte și valorile „2" indică copiii proveniți din familii cu doi părinți. Așadar este utilizată o singură coloană pentru variabila dependentă (în acest caz, emotivitatea/„Emotion") și altă coloană pentru variabila independentă (în acest caz, tipul de familie/„Family"). Deci fiecare rând în parte reprezintă un anumit copil și variabilele sale dependente și independente sunt introduse în două coloane separate în „Data Editor".

13.2. Efectuarea Testului *t* pentru eșantioane independente

Pasul 1

Selectați „Analyze", „Compare Means" și „Independent Samples T test...".

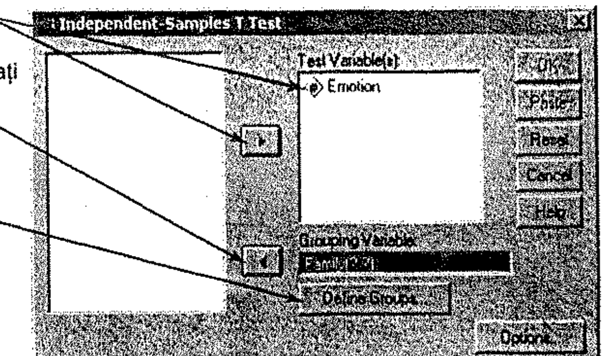


Pasul 2

Selectați „Emotion" și apoi apăsați butonul ▶ pentru a introduce această variabilă în lista variabilelor de test.

Selectați „Family" și apoi apăsați butonul ◀ pentru a introduce această variabilă în căsuța „Grouping Variable:".

Selectați „Define Groups..." pentru a defini cele două grupuri.

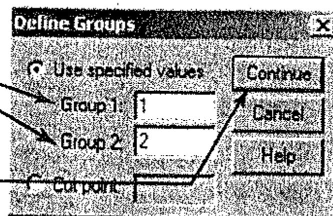


Pasul 3

Introduceţi valoarea „1" (codul pentru un singur părinte) alături de eticheta „Group1:" şi valoarea „2" (codul pentru doi părinţi) alături de eticheta „Group2:". Împărţirea grupurilor între cele două căsuţe este arbitrară.

Selectaţi „Continue".

Apăsaţi „OK" din ecranul precedent, care reapare.



13.3. Interpretarea output-ului

Group Statistics

	Family	N	Mean	Std. Deviation	Std. Error Mean
Emotion	1	10	9.50	3.100	.980
	2	12	13.42	3.370	.973

Primul tabel arată, pentru fiecare grup în parte, numărul de cazuri, media şi abaterea standard. Media pentru un singur părinte este 9,50. Există în mod evident o diferenţă între cele două tipuri de familii. Întrebarea următoare este dacă mediile diferă semnificativ.

Valoarea lui *t* este pur şi simplu media diferenţei (–3,917) divizată cu eroarea standard a diferenţei (1,382), diviziune ce produce valoarea –2,813.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Emotion	Equal variances assumed	.212	.650	–2.813	20	.011	–3.917	1.392	–6.821	–1.013
	Equal variances not assumed			–2.836	19.77	.010	–3.917	1.381	–6.800	–1.034

Dacă valoarea semnificaţiei pentru Testul Levene este mai mare de 0,05, lucru care se întâmplă aici fiind 0,650, folosiţi informaţia de pe acest prim rând. Dacă valoarea semnificaţiei pentru Testul Levene este mai mică de 0,05, folosiţi informaţia de pe rândul al doilea. Al doilea rând oferă cifrele pentru cazul în care varianţele sunt diferite semnificativ.

Pentru varianţe egale, *t* este –2,813, care cu 20 de grade de libertate este semnificativ la 0,011 pentru nivelul two-tailed. Pentru a obţine nivelul one-tailed, divizaţi cu 2 nivelul anterior şi obţineţi 0,006, rotunjit la trei zecimale.

Output-ul pentru Testul *t* pentru eşantioane independente al SPSS-ului este în mod deosebit confuz, chiar şi pentru cei cu bună cunoaştere a statisticii. Motivul este acela că există două versiuni pentru acest Test *t*. Alegerea variantei de utilizat depinde de existenţa unei diferenţe semnificative între varianţele (estimate) ale celor două grupuri de scoruri.

- Examinaţi primul tabel de la output. Acesta conţine mediile şi abaterile standard ale scorurilor pentru variabila dependentă (emotivitatea) ale celor două grupuri. Observaţi că a fost adăugată, de către calculator, o cifră adiţională la numele coloanei ce conţine variabila dependentă. Această cifră adiţională arată căruia dintre cele două grupuri îi corespund rândurile. Dacă aţi etichetat variabilele, valorile etichetelor vor fi afişate în tabel.
- Pentru copiii proveniţi din familii cu doi părinţi („family 2") valoarea medie a emotivităţii este 13,42 şi abaterea standard a valorilor emotivităţii este 3,37. Pentru copiii proveniţi din familii cu un singur părinte („family 1") valoarea medie a emotivităţii este 9,50 şi abaterea standard a valorilor emotivităţii este 3,10.
- În al doilea tabel apare linia „Levene's Test for Equality of Variance" (Testul Levene pentru egalitatea varianţelor). Dacă valoarea probabilităţii este statistic semnificativă atunci varianţele sunt inegale. În caz contrar ele sunt privite ca fiind egale.
- Testul Levene pentru egalitatea varianţelor arată în acest caz faptul că varianţele sunt egale pentru că valoarea lui *p* este 0,650, deci nu este statistic semnificativă.
- În consecinţă veţi folosi rândul pentru „Equal variances assumed". Valoarea *t*, gradele sale de libertate şi probabilitatea sunt afişate. Valoarea *t* pentru varianţe egale este -2,813, care, cu 20 de grade de libertate, are un nivel (two-tailed) de semnificaţie de exact 0,011.
- Dacă Testul Levene pentru egalitatea varianţelor ar fi fost statistic semnificativ (de exemplu, 0,05 sau mai puţin) ar fi trebuit utilizată a doua linie a tabelului care oferă valorile Testului *t* pentru varianţe inegale.

13.4. Raportarea rezultatelor

- Am putea raporta rezultatele acestei analize sub forma: „Media pentru valorile testelor de emotivitate ale copiilor ce provin din familii cu doi părinţi (M = 13,42, SD = 3,37) este semnificativ mai mare (t = –2,81, DF = 20, two-tailed p = 0,011) decât aceea a copiilor proveniţi din familii cu un singur părinte (M = 9,50, SD = 3,10)".
- Este neobişnuit să întâlniţi Testul *t* pentru varianţe inegale în rapoartele psihologice. Mulţi psihologi nu sunt conştienţi de existenţa lui. Aşadar, ce se va întâmpla dacă ar fi să folosiţi unul? Pentru a clarifica lucrurile, vom scrie: „Din cauza faptului că varianţele pentru cele două grupuri sunt diferite semnificativ (F = 8,43; p < 0,05), a fost utilizat un Test *t* pentru varianţe inegale...".
- Dacă preferaţi folosirea intervalelor de încredere, aţi putea scrie: „Diferenţa dintre valorile testelor de emotivitate ale copiilor ce provin din familii cu doi părinţi (M = 13,42, SD = 3,37) şi cei proveniţi din familii cu un singur părinte (M = 9,50, SD = 3,10) este –3,92. Intervalul de încredere 95% pentru această diferenţă este de la –6,82 la –1,01. Deoarece intervalul nu conţine punctul 0,00 diferenţa este statistic semnificativă la nivelul de semnificaţie two-tailed de 5%".

14 Testul chi-square
Diferențe între frecvențele eşantioanelor

Rezumat

- În general, Testul chi-square este folosit pentru evaluarea existenței unei diferențe semnificative între două sau mai multe eşantioane care sunt formate din date de frecvență (date nominale). Cu alte cuvinte, este testul statistic uzual care analizează tabelele de asociere sau contingență pe baza a două variabile categoriale nominale.
- De asemenea, poate fi folosit şi pentru testarea faptului că un singur eşantion este semnificativ diferit de o populație cunoscută. Această aplicație este cea mai puțin comună, doarece caracteristicile unei populații sunt rareori cunoscute în cercetare.
- Este extrem de important să rețineți faptul că Testul chi-square analizează frecvențe. Acestea nu trebuie *niciodată* să fie transformate în *procente* pentru a fi introduse în SPSS, căci vor da rezultate eronate atunci când se vor calcula valoarea şi semnificația lui chi-square. Trebuie să existe o distincție foarte clară între acest lucru şi folosirea procentelor atunci când încercăm să interpretăm ce se întâmplă în cadrul unui tabel de contingență.
- De asemenea, trebuie să aveți în vedere că o analiză chi-square trebuie să includă datele fiecărui individ o singură dată. Aşadar, frecvențele totale trebuie să fie egale cu numărul de persoane folosite în analiză.
- Analizarea şi interpretarea tabelelor de contingență 2 × 2 sunt procese directe. Totuşi, interpretarea unor tabele de contingență mai mari nu este la fel de uşoară, căci poate necesita împărțirea tabelului într-o serie de tabele mai mici. Partiționarea Testului chi-square, aşa cum este cunoscut procedeul, necesită, de obicei, adaptarea la nivelele de semnificație pentru a lua în considerare şi numărul de subanalize efectuate. Consultați un manual de statistică pentru mai multe detalii.
- Acest capitol include şi Testul Fisher, care poate fi folositor în anumite condiții atunci când datele dumneavoastră nu îndeplinesc condițiile pentru chi-square (mai ales atunci când frecvențele anticipate sunt prea mici).
- Este discutat şi Testul McNemar pentru semnificația modificărilor, care se află în strânsă legătură cu chi-square.
- Unele variante ale Testului chi-square sunt folosite ca măsuri ale parametrului *goodness-of-fit* în unele dintre cele mai avansate tehnici statistice care vor fi discutate mai târziu în această carte, cum ar fi regresia logistică. Testul pentru *goodness-of-fit* nu face decât să evalueze relația dintre datele disponibile şi datele anticipate pe baza unui set de variabile predictori. În consecință, este

foarte importantă înțelegerea adecvată a Testului chi-square, nu doar din cauza modului simplu în care se aplică, ci şi datorită rolului pe care îl îndeplineşte în cadrul tehnicilor statistice mai avansate.

Tabelul 14.1. *Relația dintre programul TV favorit şi viața sexuală*

Respondents	Soap opera	Crime drama	Neither
Males	27	14	19
Females	17	33	9

Procesarea Testului chi-square folosind două sau mai multe eşantioane este exemplificată cu datele din tabelul 14.1 (*ISP*, tabelul 14.8). Acest tabel arată care dintre cele trei tipuri de program de televiziune este preferat de către un eşantion de 119 adolescenți de ambele sexe. Pentru a analiza cu ajutorul aplicației SPSS un astfel de tabel cu date, mai întâi trebuie să introducem datele în „Data Editor" şi să ponderăm celulele după frecvențele cazurilor pe care le conțin.

- Din moment ce avem de-a face cu un tabel deja format, este nevoie să efectuăm, mai întâi, procedura „Weighting cases" (vezi secțiunea 14.1). Altfel, ați completa tabelul 14.1 caz după caz, indicând cărei categorii de pe un rând şi cărei categorii de pe o coloană îi aparține fiecare caz (vezi secțiunea 14.2). Trebuie să identificăm fiecare dintre cele şase celule din tabelul 14.1. Rândurile tabelului reprezintă sexul participanților, în timp ce coloanele reprezintă cele trei tipuri de program de televiziune. Apoi, vom pondera fiecare dintre cele şase celule ale tabelului după numărul de cazuri pe care le conțin.
- Prima coloană, numită „Sex" în pasul 1 al secțiunii 14.1, conține coduri pentru băieți (1) şi fete (2). (Şi aceste valori au fost etichetate.)
- A doua coloană, numită „Program", conține codurile celor trei tipuri de program de televiziune: telenovelă (1), dramă cu crime (2) şi nici una (3). (Şi aceste valori au fost etichetate.)

14.1. Introducerea datelor din tabelul 14.1, folosind procedura „Weighing cases"

Pasul 1

În meniul „Variable View" din „Data Editor", etichetați primele trei coloane cu „Sex", „Program" şi, respectiv, „Freq".

Îndepărtați cele două zecimale.

Etichetați valorile „Sex" şi „Program".

	Name	Type	Width	Decimals	Label	Values
1	Sex	Numeric	8	0		{1, Males}...
2	Program	Numeric	8	0		{1, Soap}...
3	Freq	Numeric	8	0		None
4						

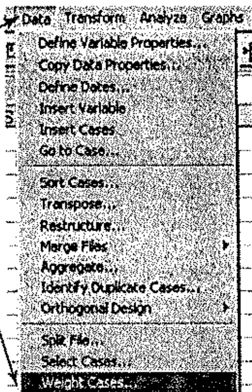
Pasul **2**

Introduceți valorile adecvate în meniul „Data View" din „Data Editor". Fiecare rând reprezintă una dintre cele șase celule din tabelul 14.1.

	Sex	Program	Freq
1	1	1	27
2	1	2	14
3	1	3	19
4	2	1	17
5	2	2	33
6	2	3	9
7			

Pasul **3**

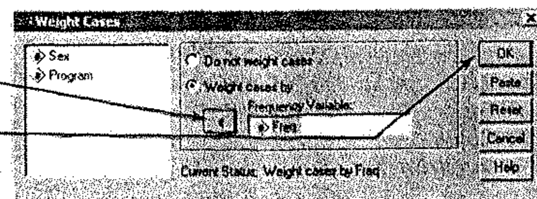
Pentru a pondera aceste celule, selectați „Data" și „Weight Cases...".



Pasul **4**

Selectați „Freq", „Weight cases by" și apoi apăsați butonul ◄.

Apăsați „OK".



14.2. Introducerea datelor din tabelul 14.1 caz cu caz

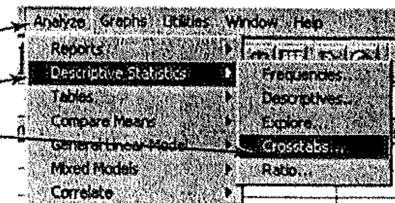
Introduceți valorile pentru cele două variabile în fiecare dintre cele 119 cazuri.

	Sex	Program
1	1	1
2	1	1
3	1	1
4	1	1

14.3. Efectuarea unui Test chi-square pe baza tabelului 14.1

Pasul **1**

Selectați „Analyze", „Descriptive Statistics" și „Crosstabs..." („Tabele de asociere...").

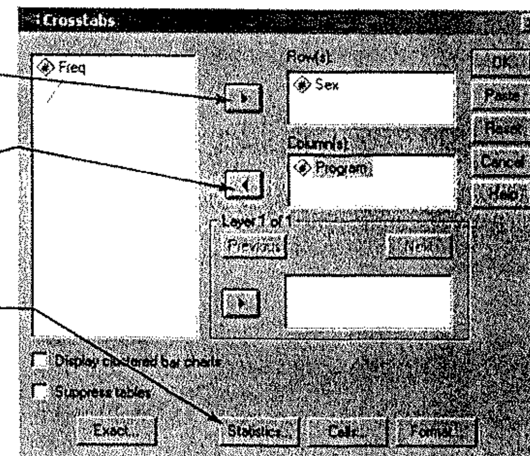


Pasul **2**

Selectați „Sex" și apăsați butonul ► pentru „Row(s):" pentru a-l introduce în respectiva casetă.

Selectați „Program" și apăsați butonul ◄ pentru „Column(s):" pentru a-l introduce în caseta respectivă.

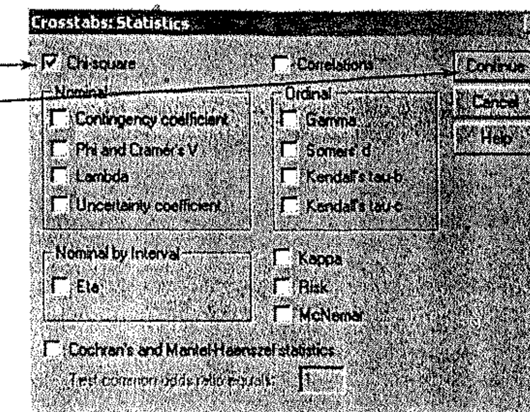
Selectați „Statistics...".



Pasul **3**

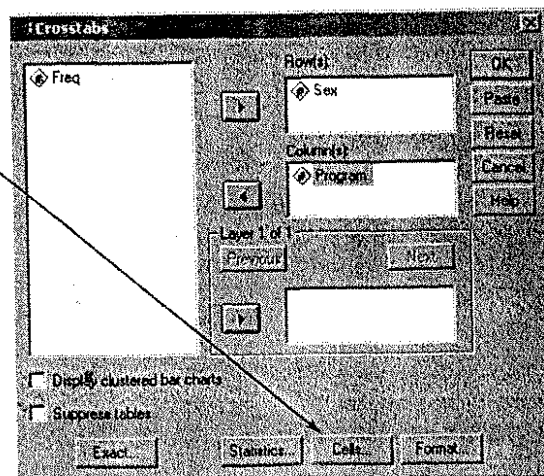
Selectați „Chi-square".

Selectați „Continue".



Pasul 4

Selectați „Cells".



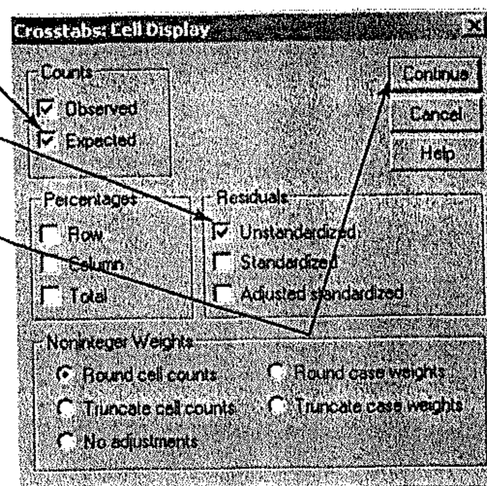
Pasul 5

În secțiunea „Counts" selectați „Expected".

Selectați „Unstandardized" în secțiunea „Residuals". (Termenul „rezidual" se referă la diferențe.)

Selectați „Continue".

Apăsați „OK" în ecranul anterior, care va reapărea.



14.4. Interpretarea output-ului pentru Testul chi-square

Acest al doilea tabel din output indică frecvența („Count"), frecvența așteptată („Expected Count") și diferența („Residual") dintre cele două pentru fiecare dintre cele șase celule ale tabelului.

Sex * Program Crosstabulation

			Program			Total
			1	2	3	
Sex	1	Count	27	14	19	60
		Expected Count	22.2	23.7	14.1	60.0
		Residual	4.8	-9.7	4.9	
	2	Count	17	33	9	59
		Expected Count	21.8	23.3	13.9	59.0
		Residual	-4.8	9.7	-4.9	
Total		Count	44	47	28	119
		Expected Count	44.0	47.0	28.0	119.0

De exemplu, frecvența sau numărul de fete care spun că preferă telenovelele este de 17, iar numărul anticipat de probabilitate este de 21,8, iar diferența dintre cele două valori este de –4,8.

Trebuie să înțelegeți ce înseamnă cifrele pentru „Count" din tabelul de asociere dintre gen și program. Acestea sunt echivalente cu tabelul inițial cu date (tabelul 14.1). Puteți transforma cifrele în procente (vedeți captura de ecran din pasul 5 al secțiunii 14.3, unde se pot vedea trei opțiuni diferite – „Row", „Column" și „Total" în meniul pentru procente). Alegerea procentajelor pe coloană v-ar ajuta, probabil, să vedeți că băieții tind să prefere telenovelele și categoria „nici una" mai mult decât fetele. Fetele tind să prefere programele cu crime mai mult decât băieții.

Al treilea și cel din urmă tabel indică valoarea coeficientului (Pearson) chi-square (13,518), gradele de libertate (2) și semnificația two-tailed (0,001). Din moment ce această valoare este mai mică de 0,05, coeficientul chi-square este semnificativ.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	13.518[a]	2	.001
Likelihood Ratio	13.841	2	.001
Linear-by-Linear Association	.000	1	.987
N of Valid Cases	119		

a. 0 celule (,0%) au frecvența anticipată mai mică de 5. Frecvența așteptată minimă este de 13,88.

- Al doilea tabel indică frecvențele observate și cele așteptate ale cazurilor și diferența (valoarea reziduală) dintre acestea pentru fiecare celulă. Frecvența observată (numită „Count") este prezentată mai întâi, iar apoi frecvența așteptată (numită „Expected Count"). Frecvențele observate sunt întotdeauna numere întregi, astfel încât ar trebui să fie ușor de localizat. Frecvențele așteptate sunt întotdeauna exprimate cu precizie de o zecimală, astfel încât se pot identifica foarte ușor. Așadar, prima celulă din tabel (băieții cărora le plac telenovelele) are o frecvență observată de 27 și o frecvență așteptate de 22,2.
- Coloana finală din tabel (etichetată „Total") conține numărul de cazuri din respectivul rând, urmat de numărul așteptat de cazuri din tabel. Așadar, primul rând are 60 de cazuri, număr care va fi întotdeauna indentic numărului de cazuri așteptate (adică 60,0).
- În mod similar, rândul final din acest tabel (etichetat „Total") prezintă mai întâi numărul de cazuri din respectiva coloană urmat de numărul așteptat de cazuri din tabel pentru coloana respectivă. Așadar, prima coloană are 44 de cazuri, număr care va fi întotdeauna egal cu numărul așteptat de cazuri (adică 44,0).
- Valoarea chi-square, gradul său de libertate și nivelul său de semnificație sunt afișate în al treilea tabel pe rândul care începe cu „Pearson", cel care a elaborat acest test. Valoarea chi-square este de 13,518, care, rotunjită la un număr cu două zecimale, devine 13,52. Gradul său de libertate este 2, iar probabilitatea two-tailed exactă este de 0,001.
- Sub acest tabel mai putem vedea și mărimea „minimum expected count" a oricărei celule din tabel, care este 13,88 pentru ultima celulă (fetele cărora nu le place nici unul dintre programe). Dacă frecvența minimă așteptată este mai mică decât 5,0, atunci trebuie să fim atenți cu folosirea Testului chi-square. Dacă aveți un tabel 2 × 2 de tip chi-square și apar frecvențe anticipate reduse, ar fi mai bine să folosiți Testul Fisher, pe care aplicația SPSS îl include în output în astfel de situații.

14.5. Raportarea output-ului pentru Testul chi-square

Există două modalități de descriere a acestor rezultate. Pentru ochiul neexperimentat, acestea pot părea foarte diferite, dar, practic, se reduc la același lucru:

- Putem descrie rezultatele în felul următor: „A existat o diferență semnificativă între frecvențele observate și cele așteptate în cazul adolescenților băieți și fete în ceea ce privește preferința lor pentru cele trei tipuri de program de televiziune ($\chi^2 = 13{,}51$, $DF = 2$, $p = 0{,}001$)".
- O altă posibilitate, la fel de exactă, ar fi următoarea: „Există o asociere semnificativă între sex și preferința pentru diferite tipuri de program de televiziune ($\chi^2 = 13{,}51$, $DF = 2$, $p = 0{,}001$)".
- În plus, trebuie să raportăm și direcția rezultatelor. O modalitate de a face acest lucru ar fi să afirmăm că: „Fetele tind mai mult decât băieții să prefere programe cu crime și le plac mai puțin telenovelele sau ambele programe".

14.6. Testul Fisher exact

Procedura chi-square procesează Testul Fisher exact pentru tabele 2 × 2 atunci când una sau mai multe dintre cele patru celule au o frecvență anticipată mai mică decât 5. Testul Fisher este procesat folosind datele din tabelul 14.2 (*ISP*, tabelul 14.15).

Tabelul 14.2. *Memoria fotografică și genul*

	Photographic memory	No photographic memory
Males	2	7
Females	4	1

14.7. Interpretarea output-ului pentru Testul Fisher exact

Sex * Memory Crosstabulation

			Memory		Total
			Photographic	Non-photographic	
Sex	Males	Count	2	7	9
		Expected Count	3.9	5.1	9.0
		Residual	-1.9	1.9	
	Females	Count	4	1	5
		Expected Count	2.1	2.9	5.0
		Residual	1.9	-1.9	
Total		Count	6	8	14
		Expected Count	6.0	8.0	14.0

Acesta este al doilea dintre cele trei tabele din output, în care se pot vedea frecvențele observate („Count") și cele așteptate („Expected Count") pentru cele patru celule.

Testele chi-square

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	4.381[b]	1	.036		
Continuity Correction[a]	2.340	1	.126		
Likelihood Ratio	4.583	1	.032		
Fisher's Exact Test				0.91	.063
Linear-by-Linear Association	4.069	1	.044		
N of Valid Cases	14				

a. Procesat numai pentru un tabel 2 × 2.
b. 3 celule (75,0%) au o frecvență anticipată mai mică de 5. Frecvența anticipată minimă este de 2,14.

Al treilea și cel din urmă tabel din output-ul SPSS. Prezintă valorile statisticilor, gradul lor de libertate și nivelele de semnificație ale acestora. Semnificația Testului Fisher exact pentru acest tabel este de 0,091 la two-tailed și de 0,063 la one-tailed.

14.8. Raportarea output-ului pentru Testul Fisher exact

- Putem scrie după cum urmează: „Nu a existat nici o relaţie semnificativă între gen şi posesia unei memorii fotografice (probabilitate two-tailed exactă Fisher $p = 0{,}091$)" sau: „Bărbaţii şi femeile nu sunt diferiţi în ceea ce priveşte frecvenţa posedării unei memorii fotografice (probabilitate two-tailed exactă Fisher $p = 0{,}091$)".
- Totuşi, având în vedere dimensiunea extrem de redusă a eşantionului, constatarea ar putea fi considerată, în cel mai bun caz, ca având o semnificaţie marginală sau ca o recomandare puternică să se efectueze studii mai aprofundate în acest sens, pentru stabilirea cu o mai mare certitudine a adevărului afirmaţiei că, în cazul fetelor, memoria fotografică este o caracteristică mai des întâlnită.

14.9. Testul chi-square pentru un singur eşantion

Procesarea unui Test chi-square pentru un singur eşantion este exemplificată folosind date din tabelul 14.3 (*ISP*, tabelul 14.18), care indică frecvenţele observate şi anticipate ale zâmbetelor la 80 de bebeluşi. Frecvenţele anticipate au fost obţinute în urma unui studiu anterior pe scară largă.

Tabelul 14.3. *Date pentru un Test chi-square pentru un singur eşantion*

	Clear smilers	Clear non-smilers	Impossible to classify
Observed frequency	35	40	5
Expected frequency	40	32	8

Pasul 1

Introduceţi datele în meniul „Data View" al „Data Editor"-ului, denumind variabilele şi îndepărtând cele două zecimale.

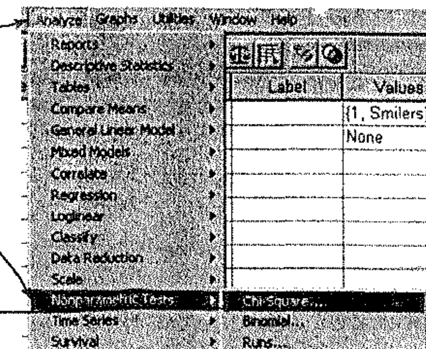
Etichetaţi cele trei categorii.

Ponderaţi celulele sau cazurile după frecvenţă („Freq").

	Category	Freq
1	1	35
2	2	40
3	3	5

Pasul 2

Selectaţi „Analyze", Nonparametric Tests" şi „Chi-square...".

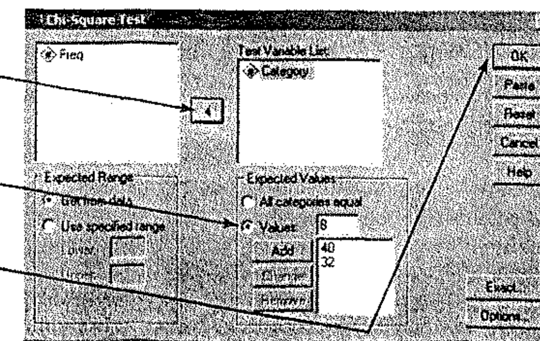


Pasul 3

Selectaţi „Category" şi apăsaţi butonul ◄ pentru a introduce parametrul în caseta „Test Variable List:".

Introduceţi pe rând valorile anticipate ale frecvenţelor în caseta „Values:", apăsând „Add" după fiecare dintre ele.

Apăsaţi „OK".



14.10. Interpretarea output-ului unui Test chi-square pentru un singur eşantion

Category

	Observed N	Expected N	Residual
Smilers	35	40.0	-5.0
Non-smilers	40	32.0	8.0
Unclassifiable	5	8.0	-3.0
Total	80		

Primul dintre cele două tabele din output indică frecvenţele observate şi pe cele anticipate pentru cele trei categorii, împreună cu diferenţa sau valoarea reziduală dintre ele. Prima coloană conţine trei categorii, a doua coloană numărul *N* de frecvenţe observate, iar cea de-a treia numărul *N* de frecvenţe aşteptate, iar în cea de-a patra putem vedea diferenţa dintre frecvenţele observate şi anticipate. Frecvenţa observată a zâmbetelor este de 35, iar cea anticipată era de 40,0.

Test Statistics

	Category
Chi-Square[a]	3.750
df	2
Asymp. Sig.	.153

a. 0 celule (0%) aveau frecvențe anticipate mai mici de 5%.

Frecvența anticipată minimă pentru fiecare celulă a fost de 8,0. Al doilea tabel arată valoarea lui chi-square (3,750), gradul de libertate (2) și nivelul de semnificație (0,153). Din moment ce nivelul de semnificație este mai mic de 0,05, frecvențele observate nu diferă semnificativ de cele anticipate datorită șansei (întâmplării).

14.11. Raportarea output-ului unui Test chi-square pentru un singur eșantion

Putem descrie rezultatele analizei după cum urmează: „Nu a existat nici o diferență statistică între frecvențele observate și anticipate pentru cele trei categorii de zâmbet la bebeluși (χ^2 = 3,75, *DF* = 2, *p* = 0,153)".

14.12. Testul McNemar

Procesarea Testului McNemar este exemplificată folosind datele din tabelul 14.4 care indică numărul de adolescenți care și-au schimbat sau nu părerea în legătură cu urmarea studiilor universitare după ce au asistat la un discurs despre carieră care favoriza educația universitară (*ISP*, tabelul 14.19). Tabelul indică numărul celor care voiau să meargă la universitate și înainte și după discurs (30), cei care voiau să meargă înainte de discurs, dar nu și după (10), cei care voiau să meargă la universitate în urma discursului, dar nu înainte (50) și cei care nu voiau să meargă la universitate nici înainte, nici după discurs (32).

Tabelul 14.4. *Studenți care vor să urmeze cursuri universitare înainte și după un discurs despre carieră*

	1 Before talk 'yes'	2 Before talk 'no'
1 After talk 'yes'	30	50
2 After talk 'no'	10	32

Pasul 1

Introduceți datele în meniul „Data View" din „Data Editor", denumind variabilele și îndepărtând cele două zecimale.

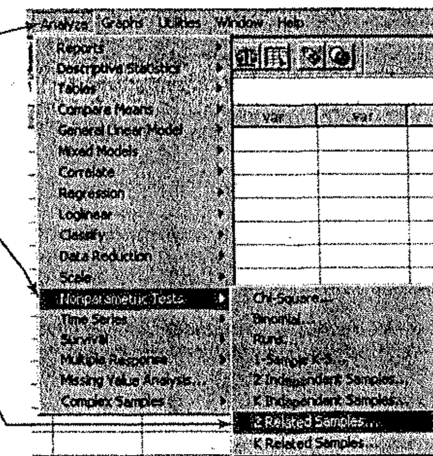
Etichetați valoarea 1 cu „Da" și valoarea 2 cu „Nu".

Ponderați celulele sau cazurile după „Freq".

	After	Before	Freq
1	1	1	30
2	1	2	50
3	2	1	10
4	2	2	32
5			

Pasul 2

Selectați „Analyze", „Nonparametric Tests" și „2 Related Samples...".



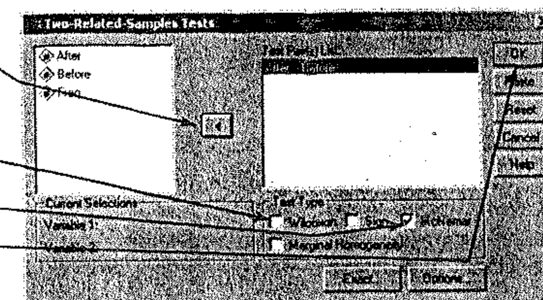
Pasul 3

Selectați „After", „Before" și apăsați butonul ◄ pentru a introduce aceste două variabile în caseta „Test Pair(s) List:".

Deselectați opțiunea „Wilcoxon".

Selectați „McNemar".

Apăsați „OK".



14.13. Interpretarea output-ului pentru Testul McNemar

Primul dintre cele două tabele prezintă frecvenţele cazurilor din cele patru celule, ca în tabelul 14.4. Cele două valori, „1" şi „2", nu au fost etichetate.

After & Before

	Before	
After	1	2
1	30	50
2	10	32

Numărul celor care şi-au schimbat părerea în sensul că iniţial voiau să urmeze cursuri universitare, iar după ce au ascultat discursul *nu* mai voiau să meargă la universitate (adică 10).

Numărul celor care şi-au schimbat părerea, iniţial nevrând să meargă la universitate, iar apoi dorindu-şi acest lucru, în urma asistării la discurs (adică 50).

Al doilea tabel din output indică numărul total N de cazuri (122), valoarea lui chi-square (23,350) şi nivelul de semnificaţie (0,000). Practic, nivelul de semnificaţie nu poate niciodată să fie zero. Valoarea sa este mai mică de 0,001. Din moment ce această valoare este mai mică decât 0,05, rezultă că este semnificativă. Acest lucru înseamnă că a existat o schimbare semnificativă la nivelul numărului de adolescenţi care şi-au schimbat părerea în legătură cu urmarea cursurilor universitare în urma asistării la o discuţie despre carieră în care se favoriza educaţia universitară.

Test Statistics[b]

	After & Before
N	122
Chi-Square[a]	25.350
Asymp. Sig.	.000

a. Continuitatea corectată
b. Testul McNemar

14.14. Raportarea output-ului pentru Testul McNemar

Putem raporta rezultatele acestei analize după cum urmează: „S-a înregistrat o creştere semnificativă la nivelul numărului de adolescenţi care s-au hotărât să urmeze cursuri universitare după ce au asistat la discuţie ($\chi^2 = 25{,}35$, $DF = 2$, $p < 0{,}001$)".

14.15. Chi-square fără tabele deja formate

În acest capitol, ne-am concentrat atenţia asupra modului în care putem analiza datele din tabele de contingenţă preexistente. Acesta este motivul pentru care avem nevoie de procedura de ponderare. Totuşi, nu veţi folosi întotdeauna tabele deja formate. Orice variabilă care este formată doar dintr-un număr mic de categorii nominale poate fi folosită pentru Testul chi-square. De exemplu, dacă doriţi să analizaţi relaţia dintre gen (codificat cu 1 pentru bărbaţi şi 2 pentru femei) şi vârstă (codificată cu 1 pentru cei cu

vârsta sub 20 de ani, 2 pentru intervalul de la 20 până la 30 de ani şi 3 pentru cei de peste 40 de ani), procedura este următoarea.

- Introduceţi codurile pentru sexe, de pildă, pentru 60 de cazuri, în prima coloană din „Data Editor".
- Introduceţi categoriile de vârstă pentru fiecare dintre aceste cazuri din rândul echivalent al următoarei coloane.
- Efectuaţi analiza chi-square. Nu mai este nevoie să urmaţi procedura de ponderare. Frecvenţele din celule vor fi calculate de SPSS.

15 Valorile lipsă

Rezumat

- Câteodată, în cercetare, se poate întâmpla să nu aveți setul complet de valori de la fiecare participant. Valorile lipsă „explică" calculatorului cum să facă față situației.
- În mod tipic, atunci când se codifică variabilele pentru a fi introduse în foile de lucru din SPSS, cercetătorul alege o valoare distinctă pentru a indica o valoare lipsă a variabilei respective. Valoarea aleasă trebuie să fie în afara intervalului de valori posibile pentru date. De obicei, numere cum ar fi 9,99 și 999 sunt folosite pentru valorile lipsă. Este posibil să aveți mai mult de o valoare lipsă pentru oricare variabilă. De exemplu, cercetătorul va dori să distingă între circumstanțele în care participantul a refuzat să răspundă la chestionar și cazurile când acesta a omis o întrebare din cu totul alt motiv.
- Valorile lipsă pot fi, de asemenea, utilizate pentru a „instrui" calculatorul să ignore cazurile cu o anumită valoare(valori) pentru o anumită variabilă.
- Trebuie să fim foarte atenți la folosirea variabilelor lipsă în SPSS. Dacă o valoare nu a fost identificată ca valoare lipsă pentru o anumită variabilă, calculatorul va analiza ceea ce a intenționat cercetătorul să fie variabilă lipsă ca pe o variabilă autentică. Acest lucru poate distorsiona serios analiza.
- Valorile lipsă pot fi folosite în două feluri principale. În ceea ce privește *listwise deletion* – cazul este șters din analiză dacă este detectată vreo valoare lipsă pentru acel caz. Acest lucru poate micșora rapid numărul de participanți la un studiu. Alternativa este de a șterge cazul din analizele care includ variabilele pentru care există valori lipsă.
- Este mai bine să specificați valorile lipsă drept o valoare decât să nu introduceți nimic pentru o intrare. Aceasta datorită faptului că unele greșeli de introducere în scris a datelor sunt ușor de confundat cu lipsa reală a valorilor.

Când colectați datele, informațiile pentru unele dintre cazuri ale unora dintre variabile ar putea lipsi. Luați, de exemplu, datele din tabelul 15.1, care consistă din notele de la muzică și matematică, împreună cu codul pentru gen și vârsta exprimată în ani. Nu există nici o informație lipsă pentru nici unul dintre cele 10 cazuri. Dar presupuneți că primele două cazuri au lipsit la muzică, deci nu avem nici o valoare pentru ele. Ar fi păcat să nu folosim nici o dată de la aceste două cazuri, pentru că avem informații pentru celelalte trei variabile: matematică, gen și vârstă. În consecință vom introduce datele pentru aceste variabile.

Tabelul 15.1 *Notele la muzică și matematică pentru 10 copii alături de categoria genului și de cea a vârstei*

Music score	Mathematics score	Sex	Age
2	8	1	10
6	3	1	9
4	9	2	12
5	7	1	8
7	2	2	11
7	3	2	13
2	9	2	7
3	8	1	10
5	6	2	9
4	7	1	11

Deși am putea lăsa celula pentru valoarea variabilei „muzică" goală pentru cele două cazuri, ceea ce se realizează de obicei este codarea datelor lipsă cu un număr care nu corespunde nici unei valori posibile a variabilei. Presupunem că valoarea pentru muzică poate varia de la 0 la 10. Putem folosi orice număr, care nu este cuprins în intervalul 0-10, pentru a semnala o valoare lipsă pentru testul la muzică. Vom utiliza numărul 11 ca fiind codul pentru valoare lipsă în coloana „muzică", deci valorile de pe primele două rânduri ale primei coloane sunt 11. Presupunem că vârsta pentru cel de-al treilea caz lipsește. Vom folosi cifra 0 ca fiind codul pentru vârsta lipsă. În continuare este nevoie să-i „explicăm" SPSS-ului cum am codat valorile lipsă. Dacă nu facem acest lucru, atunci SPSS-ul va interpreta aceste coduri ca valori reale.

Valorile lipsă pot fi folosite de asemenea pentru a determina calculatorul să ignore anumite valori ale unei variabile, pe care dorim să le excludem din analiză. Așadar, de exemplu, puteți folosi valorile lipsă în ceea ce privește chi-square pentru a ignora anumite categorii. Lăsarea unui spațiu gol într-o celulă în foaia de lucru a editorului de date va rezulta într-o oprire completă (.), fiind introdus în celulă dacă face parte dintr-o matrice activă de înregistrări. La nivelul output-ului acestea sunt identificate ca valori lipsă, dar ele ar fi mai bine privite ca valori omise. Am recomanda să nu folosiți celulele goale ca o metodă de identificare a valorilor lipsă din moment ce această metodă nu face distincție între valorile lipsă și erorile de la tastatură. În mod normal numerele mari, cum ar fi 99 sau 999, reprezintă cea mai bună cale de a identifica valorile lipsă.

15.1. Definirea valorilor lipsă

Pasul 1

Selectați fișierul de la capitolul 10 dacă l-ați salvat. Dacă nu, introduceți datele.

Modificați cele două valori ale variabilei „Music" pentru cazurile 1 și 2 în „11".

Modificați pentru cazul 3 variabila „Age" în 0.

	Music	Maths	Sex	Age
1	11	8	1	10
2	11	3	1	9
3	4	9	2	0
4	5	7	1	8
5	7	2	2	11
6	7	3	2	13
7	2	9	2	7
8	3	8	1	10
9	5	6	2	9
10	4	7	1	11

Pasul 2

În „Variable View" din „Data Editor" selectați marginea din dreapta a celulei „Music" din coloana „Missing". Va apărea un pătrat gri cu trei puncte.

Selectați „Discrete missing values" și introduceți valoarea „11" în căsuța de mai jos.

Apăsați „OK". Valoarea „11" apare în celula din coloana „Missing" corespunzătoare rândului „Music".

Repetați această procedură pentru „Age", în loc de valoarea „11" introduceți „0".

15.2. Opțiunile pairwise și listwise

Unele dintre opțiunile disponibile când aveți valori lipsă sunt ilustrate cu procedura „corelare", deși opțiuni similare sunt disponibile împreună cu alte proceduri statistice.

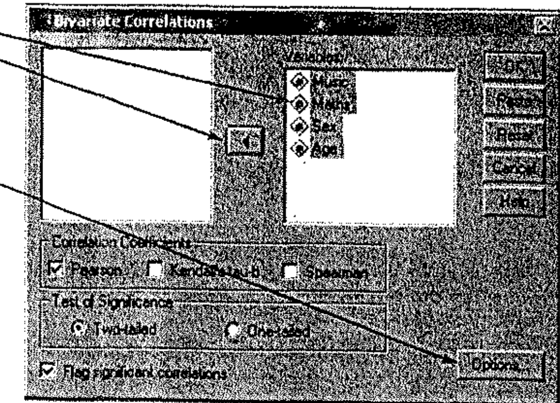
Pasul 1

Selectați „Analyze", „Correlate" și „Bivariate...".

Pasul 2

Selectați „Music", „Math", „Sex" și „Age", apoi apăsați butonul ◄ pentru a introduce aceste variabile în lista de variabile.

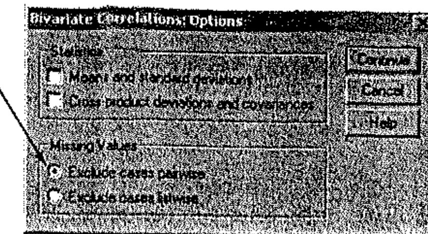
Selectați „Options...".



Pasul 3

Observați că opțiunea predefinită „Missing Values" este pairwise.
Output-ul pentru pairwise este prezentat mai jos, primul.

Dacă doriți ștergerea listwise selectați „Exclude cases listwise" și „Continue".



15.3. Eşantion din output pentru ştergerea pairwise

8 cases 7 cases

Correlations

		Music	Maths	Sex	Age
Music	Pearson Correlation	1	-.923**	.293	.681
	Sig. (2-tailed)	.	.001	.482	.092
	N	8	8	8	7
Maths	Pearson Correlation	-.923**	1	-.161	-.550
	Sig. (2-tailed)	.001	.	.656	.125
	N	8	10	10	9
Sex	Pearson Correlation	.293	-.161	1	.118
	Sig. (2-tailed)	.482	.656	.	.762
	N	8	10	10	9
Age	Pearson Correlation	.681	-.550	.118	1
	Sig. (2-tailed)	.092	.125	.762	.
	N	7	9	9	9

9 cases

10 cases

**. Correlation is significant at the 0.01 level (2-tailed).

- Ştergerea pairwise presupune că o corelaţie va fi calculată pentru toate cazurile care nu au valori lipsă, pentru orice pereche de variabile. Din moment de sunt două valori lipsă pentru testul la muzică şi nici o valoare lipsă pentru testul de matematică şi gen, numărul de cazuri pe care vor fi bazate aceste corelaţii este 8. Pentru o valoare a variabilei „vârstă" lipseşte pentru un alt caz, numărul de cazuri pe care vor fi bazate corelaţiile dintre muzică şi gen este 7. Deoarece nu este nici o valoare lipsă pentru testele de matematică şi gen, numărul de cazuri pe care vor fi bazate aceste corelaţii este 10. În cele din urmă numărul de cazuri pe care vor fi bazate corelaţiile dintre rezultatele la matematică şi vârstă este 9, pentru că există o valoare lipsă pentru vârstă şi nici una pentru matematică.
- Observaţi că numărul de cazuri variază pentru ştergerea pairwise a valorilor lipsă.

15.4. Eşantion din output pentru ştergerea listwise

- La ştergerea listwise corelaţiile sunt calculate pentru toate cazurile care nu au valori lipsă a variabilelor care au fost selectate pentru această procedură. În acest exemplu, numărul de cazuri care nu au valori lipsă pentru nici una dintre cele patru variabile este 7, ceea ce reprezintă rezultatul acestei proceduri.
- *Observaţi că numărul de cazuri nu variază pentru ştergerea totală listwise a valorilor lipsă.*

Correlations[a]

		Music	Maths	Sex	Age
Music	Pearson Correlation	1	-.956**	.354	.681
	Sig. (2-tailed)	.	.001	.437	.092
Maths	Pearson Correlation	-.956**	1	-.483	-.729
	Sig. (2-tailed)	.001	.	.272	.063
Sex	Pearson Correlation	.354	-.483	1	.088
	Sig. (2-tailed)	.437	.272	.	.852
Age	Pearson Correlation	.681	-.729	.088	1
	Sig. (2-tailed)	.092	.063	.852	.

**. Correlation is significant at the 0.01 level (2-tailed).

a. Listwise N=7

Toate cazurile sunt în număr de 7.

15.5. Interpretarea output-ului

Nu este mai nimic la output care să nu fi fost discutat în alte capitole. Sigurul lucru care merită reţinut este acela că statistica se bazează pe un număr redus de cazuri.

15.6. Raportarea output-ului

Amintiţi-vă să raportaţi mărimile veritabile ale eşantionului (sau gradele de libertate) folosite pentru raportarea fiecărei analize statistice în loc de numărul de cazuri per total.

16 Recodificarea valorilor

Rezumat

- Din când în când, cercetătorii trebuie să modifice modul în care se înregistrează pe computer anumite valori ale unei variabile; uneori, trebuie combinate mai multe valori diferite într-una singură.
- Procedura de recodificare a valorilor oferă o flexibilitate considerabilă prin permiterea efectuării unor modificări rapide şi simple ale codificării numerice a oricărei valori.
- Deoarece aplicaţia SPSS poate recodifica valori în mod rapid şi cu uşurinţă, este bine să introduceţi cât mai mulţi itemi cu date în forma lor originală. Acest lucru îi oferă cercetătorului o mare libertate de recodificare a datelor. Dacă acestea din urmă au fost înregistrate de către cercetător înainte de introducerea lor în foaia de lucru, nu se mai poate încerca o codificare alternativă a datelor originale. De exemplu, dacă toate scorurile sunt introduse pentru măsurarea extroversiunii fiecărui participant, SPSS se poate folosi pentru recodificarea datelor introduse (cum ar fi inversarea scorului unui item) sau pentru calcularea unui total pe baza tuturor itemilor de extroversiune sau a unei părţi dintre aceştia. Dacă cercetătorul calculează manual o valoare a extroversiunii înainte de a o introduce în SPSS, nu mai este posibilă refacerea scorului chestionarului într-o manieră diferită, fără să adăugăm şi scorurile originare.
- De obicei, este cel mai bine să menţinem intacte datele originare. Aşadar, trebuie să vă asiguraţi întotdeauna că aţi creat o variabilă complet nouă (o coloană nouă) pentru variabila înregistrată. *Nu modificaţi* variabilele decât în cazul în care sunteţi absolut siguri că doriţi să modificaţi permanent datele originare.

Uneori, este necesar să recodificăm valorile unei anumite variabile din datele noastre. Pot exista numeroase motive pentru acest lucru, printre care şi următoarele:

- Să grupăm câteva categorii ale unei variabile nominale care, altfel, ar avea foarte puţine cazuri. Această metodă este folosită în mod obişnuit în metode statistice de tipul chi-square.
- Să grupăm variabilele numerice în intervale de scoruri.
- Să recodificăm itemii care trebuie inversaţi (vezi capitolul 17).

Poate dorim să ne clasificăm eşantionul în două sau mai multe grupuri în funcţie de o anumită variabilă, cum ar fi vârsta sau inteligenţa. Vom ilustra recodificarea cazurilor folosind datele din tabelul 16.1 care conţine scorurile abilităţilor muzicale şi matematice obţinute de către un număr de 10 copii, însoţite de codul pentru gen şi vârsta în ani.

Tabelul 16.1. *Scorurile abilităţilor muzicale şi matematice obţinute de către un număr de 10 copii, însoţite de codul pentru gen şi vârsta în ani*

Music score	Mathematics score	Sex	Age
2	8	1	10
6	3	1	9
4	9	2	12
5	7	1	8
7	2	2	11
7	3	2	13
2	9	2	7
3	8	1	10
5	6	2	9
4	7	1	11

Scorurile abilităţilor muzicale şi matematice sunt aceleaşi pe care le-am prezentat anterior în tabelul 7.1. Să presupunem că dorim să procesăm corelaţia dintre valorile abilităţilor muzicale, pe de o parte, şi celor matematice, pe de alta, în cazul copiilor mici şi a celor mai mari. Pentru a face acest lucru trebuie să hotărâm mai întâi câte grupe de vârstă dorim. Din moment ce avem la dispoziţie doar 10 copii, ne vom mulţumi cu două grupuri. Apoi, vom decide care este punctul de separaţie a vârstei dintre cele două grupuri. Din moment ce dorim ca cele două grupuri să fie de mărime considerabilă, vom selecta valoarea 10 drept punct de separaţie, aceasta însemnând că în prima grupă vor fi incluşi cei care au mai puţin de 10 ani, iar în cea de-a doua cei care au zece sau mai mulţi ani. Acum, vom folosi aplicaţia SPSS pentru a recodifica vârsta în acest fel.

16.1. Recodificarea valorilor

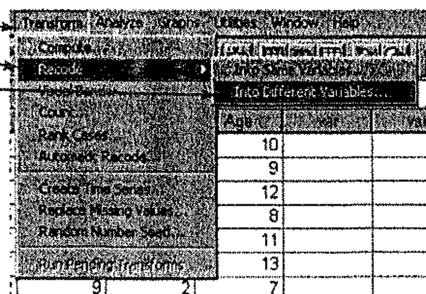
Pasul 1

Selectaţi fişierul din capitolul 10 dacă l-aţi salvat.
Dacă nu, introduceţi datele.

	Music	Maths	Sex	Age
1	2	8	1	10
2	6	3	1	9
3	4	9	2	12
4	5	7	1	8
5	7	2	2	11
6	7	3	2	13
7	2	9	2	7
8	3	8	1	10
9	5	6	2	9
10	4	7	1	11

Pasul 2

Selectați „Transform", „Recode" și „Into Different Variables...".



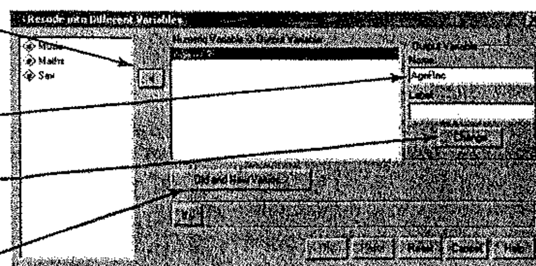
Pasul 3

Selectați „Age" și apăsați butonul ◄ pentru a introduce „Age" în caseta „Numeric Variable →Output variable:".

Introduceți numele noii variabile (de exemplu, „AgeRec") în caseta „Name:".

Selectați „Change" pentru a adăuga acest nume nou în caseta „Numeric Variable → Output Variable:".

Selectați „Old and New Values...".

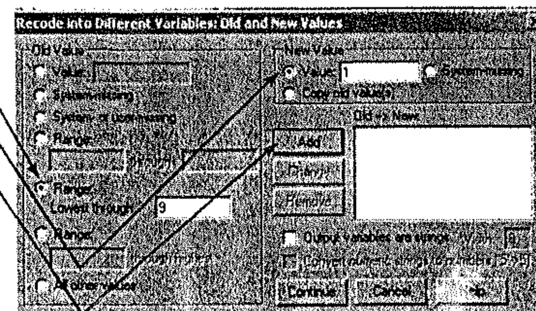


Pasul 4

Selectați „Range: Lowest through" și scrieți „9" în caseta de alături.

Selectați „Value:" din meniul „New Value" și scrieți „1" în caseta de alături.

Selectați „Add" și introduceți „Lowest through 9 →1" în caseta „Old →New".



Pasul 5

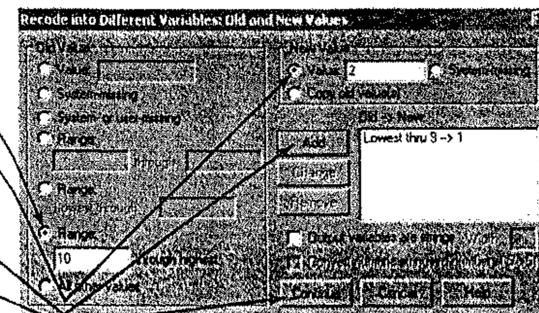
Selectați „Range: through highest" și scrieți „10" în caseta de lângă el.

Selectați „Value:" din meniul „New Value" și scrieți „2" în caseta de alături.

Selectați „Add" și introduceți „10 through Highest →2" în caseta „Old→New".

Selectați „Continue".

Apăsați „OK" în ecranul anterior, care va reapărea.



Pasul 6

Noua variabilă și valorile sale sunt afișate în meniul „Data View". Asigurați-vă că acestea sunt valorile de care aveți nevoie.

Age	AgeRec
10	2.00
9	1.00
12	2.00
8	1.00
11	2.00
13	2.00
7	1.00
10	2.00
9	1.00
11	2.00

Dacă dețineți un set complex de date, puteți uita cu ușurință ce modificări ați făcut asupra datelor dumneavoastră. Înregistrarea poate modifica radical outuput-ul unei analize efectuate cu ajutorul computerului. Trebuie să verificați cu atenție implicațiile oricăror recodificări înainte de a le raporta.

16.2. Recodificarea valorilor lipsă

Observați că atunci când există valori lipsă (ca în cazul vârstei în secțiunea 15.1, pasul 1), este nevoie să codificați aceste valori prin selectarea comenzilor „System-or user-missing" în secțiunea „Old Value" a subcasetei de dialog „Recode into Different Variables : Old and New Values" și selectarea comenzii „System-missing" din secțiunea „New Value".

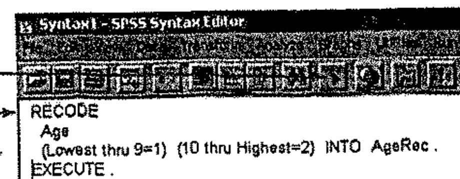
Verificați întotdeauna dacă recodificarea a funcționat așa cum trebuia comparând în editorul de date valoarea veche și cu cea nouă, pentru fiecare nouă valoare, luând unul sau mai multe cazuri.

16.3. Salvarea procedurii de recodificare ca fişier de sintaxă

Trebuie să reţineţi modul în care au fost create noile variabile. Cel mai simplu este să scrieţi. O altă metodă ar fi să salvaţi sub formă de comandă de sintaxă ceea ce aţi lucrat, folosind opţiunea „Paste" din caseta principală a procedurii. Output-ul pentru această procedură este ilustrat mai jos.

Apăsaţi butonul ▶ pentru a rula comenzile de sintaxă.

O comandă de sintaxă.



Salvaţi această comandă sub formă de fişier. Dacă doriţi, puteţi folosi fişierul pentru efectuarea acestei proceduri pe alt set de date. Înainte de elaborarea sistemului de operare Windows, comenzile din SPSS erau efectuate prin acest tip de comenzi de sintaxă.

Pentru probarea acestei proceduri, selectaţi coloana „Data View" care conţine vârsta recodificată şi ştergeţi-o. Selectaţi fereastra Syntax, selectaţi întreaga comandă existentă acolo şi rulaţi-o cu butonul indicat mai sus din bara de instrumente.

17 Calculul variabilelor noi

Rezumat

- Calculul variabilelor noi vă permite să adunaţi, să scădeţi etc. valorile câtorva variabile pentru a obţine o nouă variabilă. De exemplu, este posibil să doriţi să adunaţi câteva întrebări dintr-un chestionar, pentru a obţine un indice general al ceea ce măsoară chestionarul.
- Unul dintre puţinele dezavantaje ale SPSS-ului pentru Windows este faptul că nu se păstrează nici o înregistrare a modului în care a fost calculată variabila nouă de la bun început. Puteţi fie să vă păstraţi înregistrări scrise detaliate ale formulelor folosite pentru procesarea noii variabile, fie, dacă este suficient de simplu, să folosiţi etichetarea variabilelor pentru a descrie metoda.
- Atunci când procesaţi o nouă variabilă, în general este bine să fiţi precauţi şi să verificaţi cazurile manual de câteva ori. Este uşor să introduceţi din greşeală o formulă incorectă care va produce mai apoi o nouă variabilă care să nu fie cea pe care credeaţi că o creaţi.

Atunci când analizaţi datele, este posibil să doriţi să formaţi o nouă variabilă pe baza uneia sau mai multor variabile vechi. De exemplu, atunci când măsuraţi variabilele psihologice, se folosesc adeseori mai multe întrebări pentru a măsura mai mult sau mai puţin acelaşi lucru. De exemplu, următoarele patru afirmaţii pot fi folosite pentru evaluarea satisfacţiei faţă de viaţă:

a. În general mă bucur de viaţă.
b. Unele zile parcă mă doboară.
c. Adeseori, viaţa mi se pare destul de anostă.
d. Viitorul pare promiţător.

Participanţilor li se cere să afirme în ce măsură sunt de acord cu fiecare dintre aceste afirmaţii, pe următoarea scală de la 1 la 4:

1: Sunt întru totul de acord; 2: Sunt de acord;
3: Nu sunt de acord; 4: Îmi exprim dezacordul vehement.

Putem folosi aceşti patru itemi pentru a determina nivelul de satisfacţie în viaţă a oamenilor, prin adunarea răspunsurilor date tuturor celor patru afirmaţii.

Observaţi o problemă care apare în mod frecvent atunci când lucrăm pe baza chestionarelor: dacă răspundeţi „Sunt întru totul de acord" la primul şi al patrulea item, indicaţi faptul că vă bucuraţi de viaţă, pe când dacă răspundeţi „Îmi exprim dezacordul vehement" la al doilea şi al patrulea item, se subînţelege că nu sunteţi mulţumiţi de viaţa dumneavoastră.

Tabelul 17.1. *Valorile satisfacţiei faţă de viaţă la trei interlocutori*

	a Enjoy lifa (recode)	b Get me down (no recode)	c Dull (no recode)	d Hopeful (recode)
Respondent 1	Agree (2 recoded as 3)	Agree (2)	Strongly disagree (4)	Agree (2 recoded as 3)
Respondent 2	Disagree (3 recoded as 2)	Disagree (3)	Agree (2)	Strongly disagree (4 recoded as 1)
Respondent 3	Strongly agree (1 recoded as 4)	Disagree (3)	Disagree (3)	Disagree (3 recoded as 2)

Avem nevoie de scoruri mai mari care să indice o satisfacţie mai mare. În consecinţă, vom inversa evaluarea pentru *primul* şi *al patrulea* item, după cum urmează:

1: Îmi exprim dezacordul vehement; 2: Nu sunt de acord;
3: Sunt de acord; 4: Sunt întru totul de acord.

Putem folosi procedura de recodificare descrisă în capitolul 16 pentru a recodifica valorile primului şi celui de-al patrulea item.

Datele din tabelul 17.1 prezintă atât răspunsurile date de către trei indivizi la cele patru afirmaţii, cât şi modul în care trebuie recodificate răspunsurile la primul şi cel de-al patrulea item. Vom folosi aceste date pentru a ilustra procedura SPSS de adunare a răspunsurilor la cele patru afirmaţii, pentru a crea un indice al satisfacţiei în viaţă.

17.1. Procesarea unei variabile noi

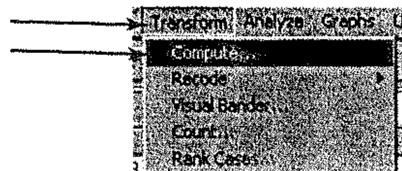
Pasul 1

Introduceţi datele (recodificate), după ce aţi denumit variabilele în felul în care este indicat şi după ce aţi îndepărtat cele două zecimale.

	q1	q2	q3	q4
1	3	2	4	3
2	2	3	2	1
3	4	3	3	2

Pasul 2

Selectaţi „Transform" şi „Compute".



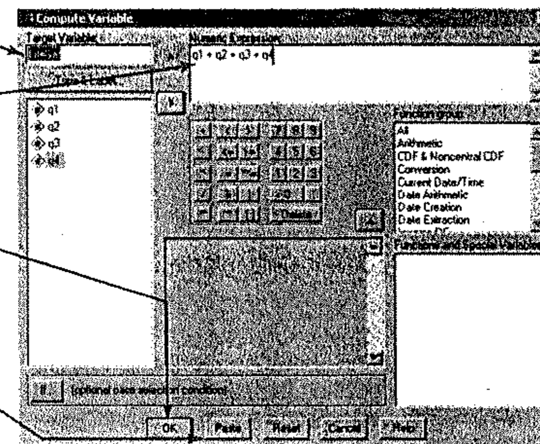
Pasul 3

Scrieţi un nume pentru noua variabilă în caseta de sub „Target Variable:" (de exemplu, „LifeSat").

Scrieţi sau selectaţi termenii expresiei şi introduceţi-i în caseta „Numeric Expression:".

Apăsaţi „OK".

Selectaţi „Paste" pentru a salva această procedură sub formă de comandă de sintaxă.



Pasul 4

Noua variabilă şi valorile sale sunt introduse în „Data Editor".

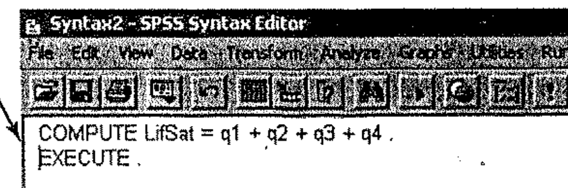
Verificaţi câteva cazuri, asigurându-vă că valorile sunt corecte.

Salvaţi fişierul dacă doriţi să-l mai folosiţi cu altă ocazie.

	q1	q2	q3	q4	LifSat
1	3	2	4	3	12.00
2	2	3	2	1	8.00
3	4	3	3	2	12.00

17.2. Salvarea procedurii de procesare ca fişier de sintaxă

Pentru a salva această procedură sub formă de fişier de sintaxă, selectaţi „Paste" în caseta principală. Această comandă de sintaxă va apărea în fereastra Syntax. (Vezi pasul 4.)



18 Testele rangurilor
Statistici nonparametrice

Rezumat

- Se poate ca uneori să doriţi să ştiţi dacă mediile a două seturi diferite de scoruri sunt semnificativ diferite una de alta, dar intuiţi că nu este îndeplinită condiţia ca scorurile fiecărei variabile să fie cât de cât normal distribuite (în formă de clopot). În aceste condiţii se pot folosi testele nonparametrice.
- Testele nonparametrice sunt teste care fac mai puţine presupuneri referitoare la caracteristicile populaţiei de la care provin datele. Această trăsătură nu este specifică testelor parametrice (cum ar fi Testul *t*) care fac mai multe presupuneri despre natura populaţiei de la care provin datele. Ipoteza normalităţii (distribuţiile de frecvenţă în formă de clopot) reprezintă un exemplu de ipoteză implicită metodelor statistice parametrice.
- Strict vorbind, metodele statistice nonparametrice nu testează diferenţele dintre medii. Acest lucru nu este posibil din moment ce acestea folosesc scoruri transformate în ranguri. De obicei, prin aceste metode se verifică dacă rangurile dintr-un grup sunt, de obicei, mai mari sau mai mici decât rangurile din celălalt grup.
- Am inclus aici Testul semnului şi Testul Wilcoxon pentru date corelate (eşantioane perechi). Cu alte cuvinte, aceste teste sunt echivalentele nonparametrice ale Testului *t* pentru eşantioane perechi. Testul Wilcoxon ar trebui să fie preferat Testului semnului atunci când se compară date numerice.
- Testul *U* Mann-Whitney se foloseşte pentru date nerelaţionate. Aşadar, acesta este echivalentul nonparametric al Testului *t* pentru eşantioane independente.

Procesarea a două teste nonparametrice pentru scoruri relaţionate este exemplificată folosind datele din tabelul 18.1, care a fost folosit şi în capitolul 12 şi care indică numărul de contacte vizuale stabilite de aceiaşi copii cu mamele lor la vârstele de 6 şi 9 luni. Observaţi că Testul semnului (secţiunea 18.1) şi Testul Wilcoxon al rangurilor pereche (secţiunea 18.4) determină nivele de semnificaţie diferite. În ceea ce priveşte detectarea diferenţelor, Testul semnului pare a fi oarecum mai puţin puternic decât Testul Wilcoxon al rangurilor pereche.

Tabelul 18.1. *Numărul de segmente de contact vizual care durează un minut, pe diferite vârste*

Baby	6 months	9 months
Clara	3	7
Martin	5	6
Sally	5	3
Angie	4	8
Trevor	3	5
Sam	7	9
Bobby	8	7
Sid	7	9

18.1. Scoruri relaţionate: Testul semnului

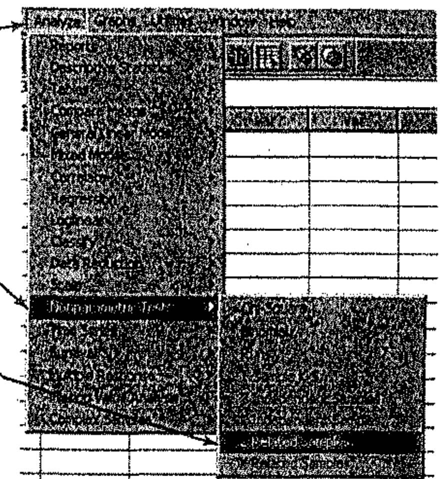
Pasul 1

Selectaţi fişierul cu date dacă l-aţi salvat. Dacă nu, introduceţi datele.

	Six mnth	Nine mnth
1	3	7
2	5	6
3	5	3
4	4	8
5	3	5
6	7	9
7	8	7
8	7	9

Pasul 2

Selectaţi „Analyze", „Nonparametric Tests" şi „2 Related Samples...".



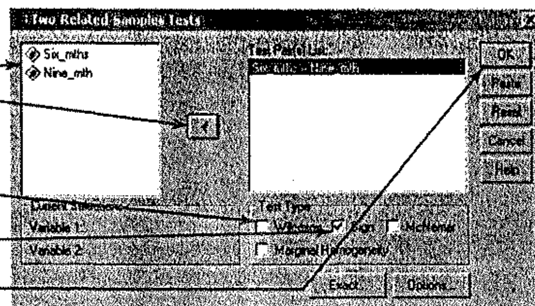
Pasul 3

Selectați „Six_mths", „Nine_mths" și apăsați butonul ◄ pentru a introduce aceste două variabile în caseta „Test Pair(s) List:".

Deselectați „Wilcoxon".

Selectați „Sign".

Apăsați „OK".



18.2. Interpretarea output-ului pentru Testul semnului

Putem ignora primul dintre cele două tabele de output. Acesta indică numărul de diferențe negative (2), pozitive (8) și inexistente (0) în ceea ce privește zâmbetul la cele două vârste.

Frequencies

		N
Nine_mth - Six_mths	Negative Differences[a]	2
	Positive Differences[b]	6
	Ties[c]	0
	Total	8

a. Nine_mth < Six_mths
b. Nine_mth > Six_mths
c. Nine_mth = Six_mths

Al doilea tabel indică nivelul de semnificație al acestui test. Probabilitatea two-tailed este de 0,289 sau de 29%, ceea ce este în mod evident nesemnificativ la nivelul de 5%. Distribuția binomială se referă la tehnica statistică prin care pot fi găsite probabilitățile în cazul eșantioanelor formate din doar două valori posibile diferite, așa cum se întâmplă în cazul Testului semnului (dat fiind faptul că vom ignora legăturile dintre valori).

Test Statistics[b]

	Nine_mth - Six_mths
Exact Sig. (2-tailed)	.289[a]

a. Am folosit distribuția binomială.
b. Testul semnului.

18.3. Raportarea output-ului pentru Testul semnului

Putem raporta aceste rezultate după cum urmează: „Nu a existat o schimbare semnificativă la nivelul numărului de contacte vizuale de la 6 la 9 luni (Testul semnului: $N = 8$, $p = 0{,}289$)".

18.4. Scoruri relaționate: Testul Wilcoxon

Testul Wilcoxon reprezintă opțiunea predefinită în caseta de dialog a testelor cu două eșantioane perechi. Dacă ați deselectat-o anterior, va trebui să o selectați din nou. Apăsați „OK" pentru a obține output-ul Testului Wilcoxon.

18.5. Interpretarea output-ului pentru Testul Wilcoxon

Putem ignora primul dintre cele două tabele de output. Acesta indică numărul de diferențe negative (2), pozitive (6) și inexistente (0) în ceea ce privește datele ordonate după cele două vârste, și media și suma datelor catalogate negative și pozitive. Valorile pentru „Nine_mths" sunt mai mari decât cele pentru „Six_mths".

Ranks

		N	Mean Rank	Sum of Ranks
Nine_mth - Six_mths	Negative Ranks	2[a]	3.00	6.00
	Positive Ranks	6[b]	5.00	30.00
	Ties	0[c]		
	Total	8		

a. Nine_mth < Six_mths
b. Nine_mth > Six_mths
c. Nine_mth = Six_mths

Al doilea tabel indică nivelul de semnificație al acestui test. În loc să folosească tabelul valorilor critice, computerul utilizează o formulă care stabilește o conexiune cu distribuția z. Scorul z este de −1,706, care are o probabilitate two-tailed de 0,088. Aceasta înseamnă că diferența dintre cele două variabile nu este semnificativă statistic la un nivel de 5%.

Test Statistics[b]

	Nine_mth - Six_mths
Z	−1.706[a]
Asymp. Sig. (2-tailed)	.088

a. Pe baza rangurilor negative.
b. Testul Wilcoxon al rangurilor însemnate

18.6. Raportarea output-ului pentru Testul Wilcoxon

Putem raporta aceste rezultate după cum urmează: „Nu s-a găsit nici o diferență semnificativă între numărul de contacte vizuale stabilite de bebelușii între 6 și 9 luni (Wilcoxon: $N = 8$, $z = 1{,}71$, two-tailed $p = 0{,}088$)".

18.7. Scoruri nerelaționate: Testul *U* Mann-Whitney

Vom ilustra procesarea unui test nonparametric pentru scoruri necorelate, folosind datele din tabelul 18.2, care indică scoruri ale emotivității obținute de 12 copii proveniți din familii cu doi părinți și de 10 copii proveniți din familii cu un singur părinte.

Tabelul 18.2. *Nivelul emotivității în familiile cu doi părinți și în cele cu unul singur*

Two-parent family X_1	Lone-parent family X_2
12	6
18	9
14	4
10	13
19	14
8	9
15	8
11	12
10	11
13	9
15	
16	

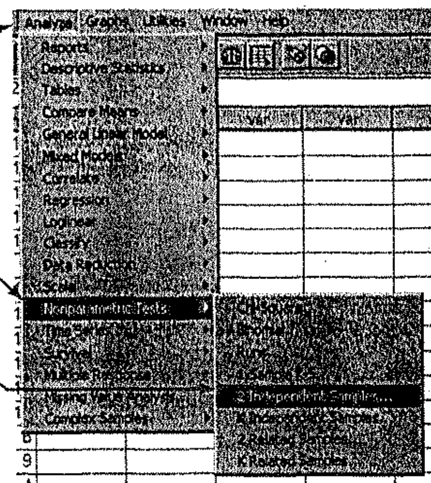
Pasul 1

Selectați fișierul cu date dacă l-ați salvat. Dacă nu, introduceți datele.

	Family	Emotion
1	2	12
2	2	18
3	2	14
4	2	10
5	2	19

Pasul 2

Selectați „Analyze", „Nonparamteric Tests" și „2 Independent Samples...".

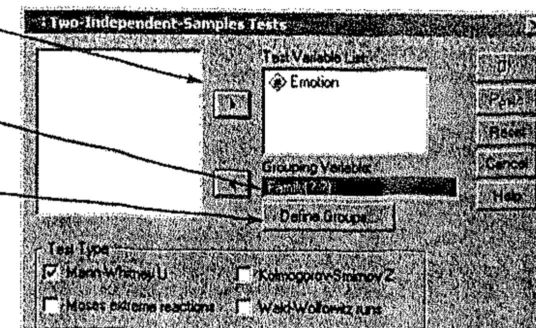


Pasul 3

Selectați „Emotion" și apăsați butonul ► pentru a introduce parametrul „Emotion" în caseta „Test Variable List:".

Selectați „Family" și apăsați butonul ◄ pentru a introduce „Family" în caseta „Grouping Variable:".

Selectați „Define Groups...".



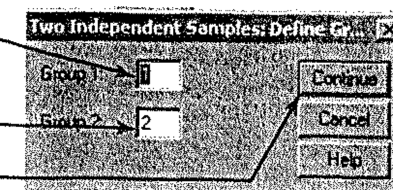
Pasul 4

Scrieți „1" (pentru un singur părinte) în caseta de lângă „Group 1:".

Scrieți „2" (pentru doi părinți) în caseta de lângă „Group 2:".

Selectați „Continue".

Apăsați „OK" în ecranul anterior, care va reapărea.



18.8. Interpretarea output-ului pentru Testul *U* Mann-Whitney

Putem ignora primul dintre cele două tabele de output. Acesta indică faptul că rangul mediu dat parametrului „Emotion" pentru primul grup (adică valoarea = 1) este de 7,85, iar rangul mediu pentru al doilea grup (adică valoare = 2) este de 14,54. Aceasta înseamnă că valorile din grupul 2 au tendința să fie mai mari decât cele din grupul 1.

Ranks

	Family	N	Mean Rank	Sum of Ranks
Emotion	1	10	7.85	78.50
	2	12	14.54	174.50
	Total	22		

Al doilea tabel indică statistica de bază Mann-Whitney, valoarea *U* fiind de 23,500, ceea ce este semnificativ din punct de vedere statistic la nivel de 0,014.

În plus, computerul a listat un scor *z* de –2,414, care este semnificativă la nivel de 0,016. Aceasta este valoarea Testului Mann-Whitney atunci când se aplică o corecție pentru ranguri înrudite. După cum se poate vedea, aceasta a modificat nivelul de semnificație doar marginal, de la 0,014 la 0,016.

Test Statistics[b]

	Emotion
Mann-Whitney U	23.500
Wilcoxon W	78.500
Z	-2.414
Asymp. Sig. (2-tailed)	.016
Exact Sig. [2*(1-tailed Sig.)]	.014[a]

a. Necorectată pentru ranguri înrudite.
b. Variabila de grupare: Familia.

18.9. Raportarea output-ului pentru Testul *U* Mann-Whitney

Putem raporta rezultatele acestei analize după cum urmează: „Testul *U* Mann-Whitney a concluzionat că scorurile emotivității la copiii din familiile cu doi părinți sunt semnificativ mai mari decât la cei din familiile cu un singur părinte ($U = 23{,}5$, $N_1 = 10$, $N_2 = 12$, p two-tailed $p = 0{,}016$)".

19 Testul raportului de varianță
Folosirea Raportului *F* pentru compararea a două variante

Rezumat

- Testul raportului de varianță (Testul *F*) indică existența unei diferențe pe care o pot manifesta două seturi de scoruri necorelate la nivelul variabilității scorurilor în jurul mediei (adică dacă variantele sunt sau nu semnificative din punct de vedere statistic).
- Acest lucru este în mod evident diferit de calcularea măsurii în care două medii sunt diferite și trebuie să țineți minte că variantele pot fi semnificativ diferite, chiar dacă mediile celor două grupuri de scoruri sunt aceleași. În consecință, examinarea varianțelor variabilelor poate fi la fel de importantă cum este compararea mediilor.
- Din cauza faptului că există puține întrebări provenite din cercetare care să fie formulate din punctul de vedere al diferențelor de varianță, cercetătorii au tendința să scape din vedere efectele asupra varianțelor și să se concentreze asupra diferențelor dintre mediile eșantioanelor. Această tendință trebuie evitată cât mai mult.
- Testul *F* este, probabil, mai des asociat cu Testul *t* și cu analiza de varianță.

Pentru a procesa raportul de varianță – sau Raportul *F* – vom împărți varianța estimată mai mare la varianța estimată mai mică. Varianța estimată se obține prin procedura „Descriptives" introdusă pentru prima dată în capitolul 5. Procesarea raportului de varianță este ilustrată folosind datele din tabelul 19.1 (*ISP*, tabelul 19.2), care raportează scorurile stabilității emoționale a pacienților cărora li s-au aplicat șocuri electrice

Tabelul 19.1. *Scorurile stabilității emoționale dintr-un studiu asupra efectelor terapiei cu șocuri electrice asupra emisferelor cerebrale*

Left hemisphere	Right hemisphere
20	36
14	28
18	4
22	18
13	2
15	22
9	1
Mean = 15.9	Mean = 15.9

fie prin emisfera dreaptă, fie prin cea stângă a creierului. În acest capitol, metoda necesită puțin calcul manual. Totuși, vă va oferi și experiență suplimenară în folosirea SPSS-ului.

O modalitatea alternativă de atingere a aceluiași obiectiv este urmărirea procedurilor Testului *t* din capitolul 13. Poate vă amintiți că Testul Levene al Raportului *F* face parte din output-ul pentru acel Test *t*. Deși Testul Levene este puțin diferit, acesta reprezintă o alternativă convențională utilă a Testului *F*.

19.1. Varianța estimată

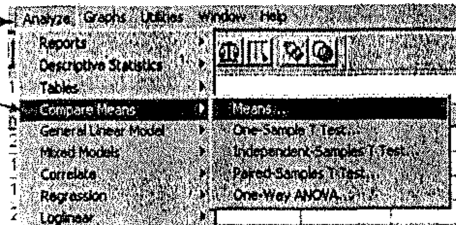
Pasul 1

Introduceți datele. Codificați emisfera stângă cu „1" și pe cea dreaptă cu „2". Etichetați aceste două valori. Numiți această variabilă prescurtat „Spheres".

	Spheres	Emotion
1	1	20
2	1	14
3	1	18
4	1	22
5	1	13
6	1	15
7	1	9
8	2	36
9	2	28
10	2	4
11	2	18
12	2	2
13	2	22
14	2	1

Pasul 2

Selectați „Analyze", „Compare Means" și „Means...".

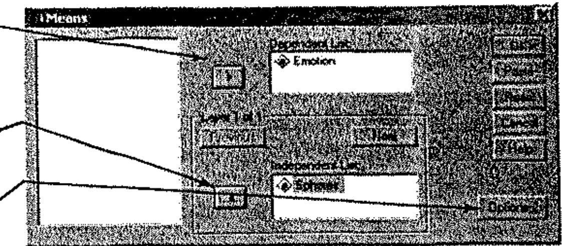


Pasul 3

Selectați „Emotion" și apăsați butonul ▶ de lângă caseta „Dependent List:" pentru a introduce parametrul acolo.

Selectați „Spheres" și apăsați pe butonul ◀ de lângă caseta „Independet List:" pentru a introduce parametrul acolo.

Selectați „Options...".

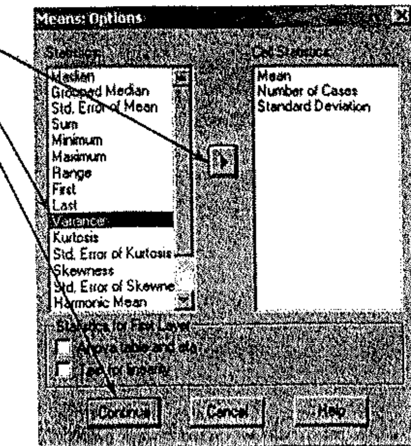


Pasul 4

Selectați „Variance" și apăsați butonul ▶ pentru a o introduce în caseta „Cell Statistics".

Selectați „Continue".

Apăsați „OK" în ecranul anterior, care va reapărea.



Report

Emotion

Spheres	Mean	N	Std. Deviation	Variance
Left	15.86	7	4.451	19.810
Right	15.86	7	13.837	191.476
Total	15.86	14	9.875	97.516

Varianțele pentru cele două grupuri se găsesc în ultima coloană a acestui tabel. Valoarea acesteia este de 19,810 pentru emisfera stângă a membrilor grupului.

19.2. Calcularea raportului de varianță din output

- Împărțiți varianța estimativă mai mare din output la varianța estimativă mai mică. Varianța estimativă mai mare este de 191,476 (pentru „Dreapta"), și, atunci când este împărțită la varianța estimată de 19,810 (pentru „Stânga") dă naștere unei varianțe sau unui Raport *F* de 9,6656. Această medie este de 9,66 dacă o rotunjim la un număr cu numai două zecimale.
- Căutați semnificația statistică a acestui raport într-un tabel de valori critice ale Rapoartelor *F*, în care gradul de libertate al raportului dintre numărător (191,48) și numitor (19,81) este 6 în ambele cazuri.
- Valoarea critică de 0,05 a Raportului *F* cu 6 grade de libertate la nivelul numărătorului și numitorului este 4,28.
- Raportul *F* pe care l-am obținut este de 9,66, care este mai mare decât valoarea critică 0,05 de 4,28 (vezi *ISP*, tabelul de semnificație 19.1, unde cea mai apropiată valoare critică este 4,4, având 5 grade de libertate la numărător).

19.3. Raportarea valorii raportului de varianță

Putem raporta aceste constatări după cum urmează: „Varianța scorurilor emotivității la pacienți cu afecțiuni în zona emisferei drepte a fost semnificativ mai mare decât cea a pacienților cu afecțiuni în zona emisferei stângi ($F_{6,6} = 9{,}66$, $p < 0{,}05$)".

20 Analiza de varianță (ANOVA) Introducere în metoda one-way ANOVA cu scoruri nerelaționate sau necorelate

Rezumat

- Analiza de varianță cu scoruri relaționate/necorelate indică măsura în care câteva (două sau mai multe grupuri) au medii foarte diferite. Această analiză presupune că fiecare dintre grupurile de scoruri provine de la indivizi diferiți. Nu este o condiție esențială să avem scoruri egale în fiecare set.
- Grupurile diferite aparțin variabilei independente. Valorile numerice corespund variabilei dependente.
- Practic, analiza de varianță calculează variația dintre scoruri și pe cea dintre mediile pe eșantioane. Ambele variații pot fi folosite pentru estimarea variației la nivelul populației. Dacă două valori estimative sunt foarte diferite, înseamnă că variația datorată variabilei independente este mai mare decât ne-am putea aștepta pe baza variației dintre scoruri. Dacă această disparitate este suficient de mare, diferența de la nivelul variabilității este semnificativă din punct de vedere statistic. Acest lucru înseamnă că variabila independentă are efect asupra scorurilor.
- Analiza de varianță poate fi dificil de interpretat atunci când se folosesc două sau mai mult de două grupuri. Analiza generală a varianței trebuie să fie semnificativă din punct de vedere statistic, dar e greu de știut care dintre cele trei sau mai multe grupuri este semnificativ diferit de celelalte.
- Soluția este descompunerea analizei în câteva comparații separate, pentru a evalua care dintre seturile de scoruri sunt semnificativ diferite de celelalte seturi de scoruri. Așadar, care dintre grupuri sunt semnificativ diferite de celelalte.
- În mod ideal, trebuie să se facă o ajustare a numărului de comparații efectuate (vezi capitolul 23 despre comparațiile multiple, pentru informații referitoare la metode mai bune decât cele descrise în aceste capitol pentru realizarea acestui lucru). Această ajustare este necesară, deoarece, cu cât sunt efectuate mai multe comparații statistice, cu atât este mai mare probabilitatea ca unele dintre comparații să fie semnificative din punct de vedere statistic.

Procesarea analizei one-way de varianță cu scoruri nerelaționate este exemplificată folosind datele din tabelul 20.1 (*ISP*, tabelul 20.2), care indică rezultatele obținute de către diferiți participanți în condiții diferite. Este vorba despre un studiu asupra efectului unor diferite tratamente hormonale și placebo asupra depresiei. Așadar, medicamentul este variabila independentă, iar depresia – variabila dependentă.

Tabelul 20.1. *Date pentru un studiu asupra efectelor hormonilor*

Group 1 Hormone 1	Group 2 Hormone 2	Group 3 Placebo Control
9	4	3
12	2	6
8	5	3

20.1. Metoda one-way ANOVA pentru eşantioane independente

Pasul 1

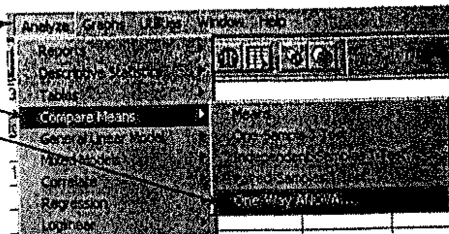
Introduceţi datele. Codificaţi cele trei condiţii cu valorile „1", „2" şi, respectiv, „3". Etichetaţi-le „Hormon 1", „Hormon 2" şi „Control placebo".

Salvaţi acest fişier pentru a-l folosi şi în capitolele 23 şi 24.

	Condition	Depression
1	1	9
2	1	12
3	1	8
4	2	4
5	2	2
6	2	5
7	3	3
8	3	6
9	3	3
10		

Pasul 2

Selectaţi „Analyze", „Compare Means" şi „One-Way ANOVA...".

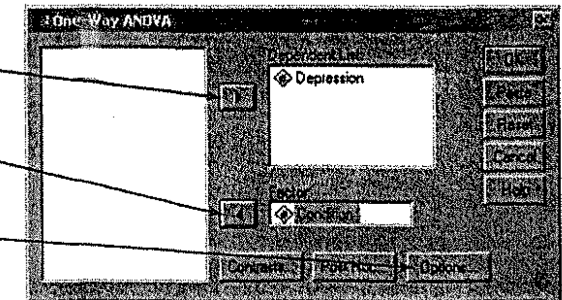


Pasul 3

Selectaţi „Depression" şi apăsaţi butonul ► de lângă caseta „Dependent List:" pentru a introduce parametrul în casetă.

Selectaţi „Condition" şi apăsaţi butonul ◄ de lângă caseta „Factor:" pentru a introduce parametrul acolo.

Selectaţi „Options...".

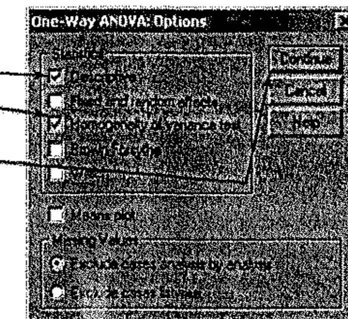


Pasul 4

Selectaţi metodele statistice „Descriptive" şi „Homogeneity of variance test".

Selectaţi „Continue".

Apăsaţi „OK" în ecranul anterior, care va reapărea.



20.2. Interpretarea output-ului

Descriptives

Depression

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
Hormone 1	3	9.67	2.082	1.202	4.50	14.84	8	12
Hormone 2	3	3.67	1.528	.882	-.13	7.46	2	5
Placebo control	3	4.00	1.732	1.000	-.30	8.30	3	6
Total	9	5.78	3.308	1.103	3.23	8.32	2	12

Primul tabel oferă diferite statistici descriptive, cum ar fi numărul de cazuri, media şi abaterea standard în cele trei condiţii şi pe eşantionul total.

Test of Homogeneity of Variances

Depression

Levene Statistic	df1	df2	Sig.
.293	2	6	.756

Al doilea tabel ne oferă rezultatele Testului Levene al similarității varianțelor. Din moment ce acest test nu este semnificativ (având o semnificație de 0,756), varianțele sunt similare sau omogene. Dacă varianțele nu ar fi omogene, ar trebui să încercăm să transformăm scorurile și să le facem să fie omogene. Altfel, pot apărea probleme la interpretarea analizei de varianță.

ANOVA

Depression

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	68.222	2	34.111	10.586	.011
Within Groups	19.333	6	3.222		
Total	87.556	8			

Al treilea tabel indică rezultatele analizei de varianță. Raportul *F* este semnificativ la nivel de 0,011, din moment ce este mai mic de 0,05.

- Raportul *F* se calculează prin împărțirea „sum of squares" (suma pătratelor abaterilor de la medie) dintre grupe este împărțită suma pătratelor de la media din interiorul grupelor („sum of squares on three groups"), ceea ce ne dă un Raport *F* de 10,586 (34,111/3,222 = 10,5869).
- Aceasta înseamnă că există o diferență semnificativă între cele trei grupuri. *Totuși, nu implică neapărat faptul că toate mediile sunt semnificativ diferite una de alta. În acest caz, bănuim că mediile de 3,67 și 4,00 nu sunt semnificativ diferite.*
- Mediile care diferă de celelalte pot fi determinate în continuare prin folosirea testelor prin comparații multiple, cum este Testul *t* pentru eșantioane independente. Pentru a face acest lucru, urmați procedura indicată pentru Testul *t* pentru eșantioane independente, așa cum este descrisă în capitolul 13. Nu trebuie să reintroduceți datele. Totuși, puteți realiza un Test *t* definind grupurile ca 1 și 2, după care le puteți redefini ca 1 și 3 și, în cele din urmă, să le redefiniți ca 2 și 3. În exemplul nostru, grupul 1 este semnificativ diferit de grupurile 2 și 3, care nu diferă semnificativ între ele. (Pentru mai multe detalii vezi *ISP*, capitolul 13.)
- Din cauza faptului că efectuăm trei comparații, nivelul exact de semnificație pentru fiecare Test *t* trebuie multiplicat de 3 ori pentru a se obține nivelul de semnificație Bonferroni.

20.3. Raportarea output-ului

Putem raporta rezultatele după cum urmează: „Per total, efectul tratamentului cu medicamente a fost semnificativ ($F_{2,6}$ = 10,58, p = 0,011). Când s-a efectuat o ajustare Bonferroni a numărului de comparații, singura diferență semnificativă a fost aceea dintre mediile tratamentului hormonal 1 și a tratamentului hormonal 2 (t = 4,02, DF = 4, two-tailed p < 0,05). Media tratamentului hormonal 1 (M = 9,67, SD = 2,08) a fost semnificativ mai mare decât aceea pentru tratamentul hormonal 2 (M = 3,67, SD = 1,53). Nu a existat nici o diferență semnificativă între media controlului placebo și media oricăruia dintre tratamentele hormonale 1 și 2".

21 Analiza de varianță pentru scoruri corelate sau măsurători repetate

Rezumat

- Analiza de varianță indică măsura în care câteva (două sau mai multe) seturi de scoruri au medii foarte diferite. Totuși, această analiză presupune că un singur eșantion de indivizi a contribuit cu scoruri pentru fiecare dintre seturile diferite de valori și se bazează pe coeficienți de corelație presupuși a fi mari între seturile de scoruri.
- Dacă datele dumneavoastră nu îndeplinesc aceste condiții, reveniți la capitolul 20 care se referă la analiza de varianță pentru eșantioane independente.
- Modificările scorurilor unei variabile în timp reprezintă un exemplu caracteristic al unui tip de studiu adecvat analizei de varianță pentru eșantioane perechi.
- Dacă este folosit corect, modelul poate fi extrem de eficient, pentru că sunt necesari mai puțini participanți pentru efectuarea studiului. Motivul este acela că, odată ce participanții sunt măsurați de mai multe ori, devine posibilă estimarea componentei diferențelor individuale a variației datelor. Într-un studiu asupra memoriei, de exemplu, unii participanți vor avea tendința să aibă rezultate bune indiferent de condiții, iar ceilalți vor avea tendința să obțină rezultate proaste. Aceste diferențe individuale pot fi identificate și ajustate în cadrul analizei. Ceea ce ar putea fi clasificat ca variație a erorii într-o analiză de varianță cu eșantioane independente este separat în două componente – componenta diferențelor individuale (eroarea implicită subiecților) și eroarea reziduală. Practic, acest lucru înseamnă că termenul „eroare" a fost redus din cauza îndepărtării componentei diferențelor individuale. Din moment ce termenul „eroare" este mai redus, putem obține rezultate semnificative folosind un număr mai mic de participanți decât în cazul unui model cu scoruri nerelaționate.

Procesarea analizei de varianță cu scoruri unifactoriale corelate este exemplificată folosind datele din tabelul 21.1, care indică valorile obținute de aceiași participanți în trei condiții diferite (*ISP*, tabelul 21.10).

Tabelul 21.1. *Valorile de ameliorare a durerii în cazul unui experiment cu medicamente*

	Aspirin	Product X	Placebo
Bob Robertson	7	8	6
Mavis Fletcher	5	10	3
Bob Polansky	6	6	4
Ann Harrison	9	9	2
Bert Entwistle	3	7	5

21.1. Medota one-way ANOVA pentru scoruri relaționate

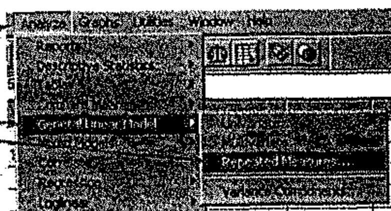
Pasul 1

Introduceți datele. Din moment ce datele sunt corelate, valorile pentru cele trei condiții se regăsesc în cele trei coloane.

	Aspirin	ProductX	Placebo
1	7	8	6
2	5	10	3
3	6	6	4
4	9	9	2
5	3	7	5

Pasul 2

Selectați „Analyze", „General Linear Model" și „Repeated Measures...".

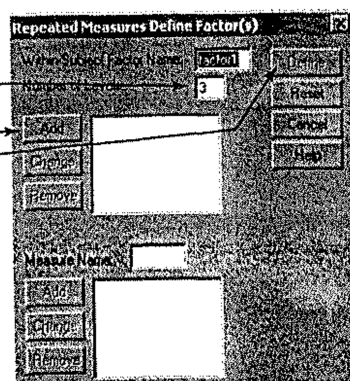


Pasul 3

Scrieți „3" în caseta „Number of Levels:".

Selectați „Add".

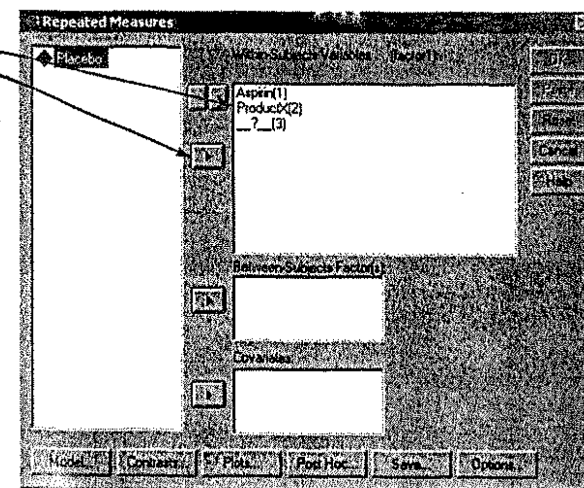
Selectați „Define".



Pasul 4

Selectați fiecare variabilă individual sau pe toate trei odată și apăsați butonul ▶ de lângă caseta „Within-Subjects Variables:" pentru a le introduce acolo.

Selectați „Options...".

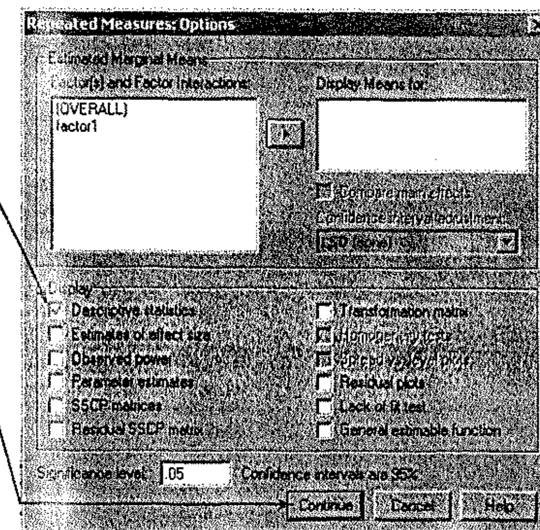


Pasul 5

Selectați „Descriptive statistics".

Selectați „Continue".

Apăsați „OK" în ecranul anterior, care va reapărea.



21.2. Interpretarea output-ului

Output-ul redă șapte tabele. În această secțiune sunt prezentate doar cele mai importante dintre ele.

Descriptive Statistics

	Mean	Std. Deviation	N
Aspirin	6.00	2.236	5
ProductX	8.00	1.581	5
Placebo	4.00	1.581	5

Al doilea tabel indică abaterile medie și standard pentru cele trei grupuri.

Mauchly's Test of Sphericity[b]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
factor1	.862	.444	2	.801	.879	1.000	.500

Testează ipoteza de nul conform căreia matricea de eroare a covarianței pentru variabilele dependente transformate ortonormalizate este proporțională unei matrici de identitate.

a. Poate fi folosit pentru ajustarea gradului de libertate în cazul testelor obișnuite de semnificație. Testele corectate sunt prezentate în tabelul „Tests of Within-Subjects Effects" ale efectelor implicite subiecților.

b. Model: intercepție.
Model implicit subiecților: factorul 1.

Cel de-al patrulea tabel indică rezultatele obținute în urma Testului Mauchly de sfericitate. Din moment ce acest test nu este semnificativ, sfericitatea este o condiție asumată și nu mai este nevoie să ajustăm nivelele de semnificație a analizei. Dacă aceasta este semnificativă, atunci trebuie folosit unul dintre testele alternative din următoarea casetă de output.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
factor1	Sphericity Assumed	40.000	2	20.000	5.106	.037
	Greenhouse-Geisser	40.000	1.758	22.752	5.106	.045
	Huynh-Feldt	40.000	2.000	20.000	5.106	.037
	Lower-bound	40.000	1.000	40.000	5.106	.087
Error(factor1)	Sphericity Assumed	31.333	8	3.917		
	Greenhouse-Geisser	31.333	7.032	4.456		
	Huynh-Feldt	31.333	8.000	3.917		
	Lower-bound	31.333	4.000	7.833		

Este necesară o parte redusă din output. În mare parte, acesta este alcătuit din analize similare folosind teste ușor diferite. Acestea sunt folosite atunci când Testul Mauchly din tabelul anterior de output este semnificativ.

Al cincilea tabel indică semnificația Raportului *F*. Aceasta este de 0,037 în condițiile în care sfericitatea este asumată.

- Raportul *F* este media pătrată (MS) a factorului 1 (20,000) împărțită la media pătrată a Erorii (factorul 1) (3,917). Acesta are valoarea de 5,106 (20,000/3,917 = 5,1059).
- Nivelul exact de semnificație a acestui Raport *F* este de 0,037. Din moment ce această valoare este mai mică de 0,05, putem trage concluzia că, per total, există o diferență semnificativă a scorurilor medii în cele trei condiții.
- Pentru a interpreta ce înseamnă analiza de varianță în cazul datelor dumneavoastră, trebuie să aveți în vedere media fiecăruia dintre cele trei grupuri de scoruri care sunt afișate în al doilea tabel. Acestea sunt 6,00, 8,00 și 4,00. Care dintre aceste medii sunt semnificativ diferite de celelalte?
- De asemenea, trebuie să țineți minte că, în cazul în care aveți trei sau mai multe grupuri, trebuie să verificați unde se regăsesc diferențele semnificative dintre perechile de grupuri. Procedura Testului *t* relaționat descrisă în capitolul 12 explică acest lucru. În exemplul de față, doar diferența dintre mediile produsului X și a metodei placebo a fost semnificativă. Din cauza faptului că efectuați mai multe Teste *t*, fiecare probabilitate exactă pentru testele t trebuie să fie înmulțită cu numărul Testului *t* care se efectuează. În exemplul nostru, există trei comparații, astfel încât fiecare probabilitate exactă trebuie să fie multiplicată cu 3. Acest procedeu este cunoscut sub numele de ajustarea sau corecția Bonferroni (vezi *ISP*, capitolul 23).

21.3. Raportarea output-ului

- Putem descrie rezultatele acestei analize după cum urmează: „O analiză de varianță one-way pentru scoruri relaționate a indicat un efect semnificativ al tratamentului în cele trei condiții ($F_{2,8} = 5,10$, $p = 0,037$). Media în cazul folosirii aspirinei a fost de 6,00, media pentru produsul X a fost 8,00, iar media placebo a fost 4,00. Nici unul dintre cele trei tratamente nu a diferit față de celelalte în cazul Testelor *t* relaționate, atunci când s-a făcut corecția Bonferroni pentru respectivul număr de comparații".
- Această analiză ar putea fi însoțită de un tabel de rezumare a analizei de varianță, așa cum este tabelul 21.2. Medicamentele reprezintă „factorul 1" din output, iar eroarea reziduală este „Error (factor 1)" din al cincilea tabel din output.

Tabelul 21.2. *Tabelul sumar al analizei de varianță*

Source of variation	Sum of squares	Degrees of freedom	Mean square	F-ratio
Drugs	40.00	2	20.00	5.11*
Residual error	31.33	8	3.92	

* Significant at 5% level.

22 Analiza de varianță bifactorială pentru scoruri nerelaționate/necorelate

Rezumat

- Analiza two-way de varianță vă permite să comparați mediile unei variabile dependente atunci când există *două* variabile independente.
- Dacă aveți mai mult de o variabilă *dependentă*, atunci nu trebuie decât să repetați analiza separat pentru fiecare variabilă. Pe de altă parte, dacă diferitele variabile dependente măsoară mai mult sau mai puțin același lucru, atunci acestea pot fi combinate într-o singură măsură generală folosindu-se procedurile de însumare descrise în capitolul 17.
- Cu aplicația SPSS, *nu* aveți nevoie de un număr egal de valori în fiecare condiție a variabilei independente. Dacă totuși reușiți să aveți un număr egal de numere în fiecare condiție, analiza va fi optimă din punct de vedere statistic.
- Deși analiza de varianță two-way ANOVA poate fi considerată un model eficient – deoarece permite încorporarea în studiu a două variabile independente –, capacitatea sa de a identifica interacțiunile poate fi și ea importantă. O interacțiune nu este decât o situație în care efectul combinat a două variabile este mai mare decât suma efectelor fiecărei variabile care acționează separat.
- Metoda two-way ANOVA poate fi dificil de interpretat. Este important să ne concentrăm asupra mediei din fiecare condiție, și nu doar asupra complexităților output-ului statistic. Este important de observat că analizele varianțelor se desfășoară conform câtorva reguli. Efectele principale sunt identificate anterior interacțiunilor. Uneori, dacă nu avem grijă, interacțiunea este confundată cu efectul principal – pur și simplu deoarece se reclamă variația pentru efectul principal înainte de a se reclama pentru interacțiune. Ca în cazul majorității analizelor statistice, este important să ne concentrăm cât mai mult asupra modelelor de medii din date drept probabilități statistice.

Realizarea unei analize two-way pentru scoruri nerelaționate de varianță este exemplificată folosind datele din tabelul 22.1. Tabelul indică scorurile obținute de către diferiți participanți în șase condiții, reflectând cei doi factori ai privării de somn și alcoolul (*ISP*, tabelul 22.11). Scopul analizei este acela de a aprecia dacă diferitele combinații de alcool și privare de somn afectează diferențiat numărul mediu de greșeli făcute.

Tabelul 22.1. *Date pentru un experiment referitor la privarea de somn : numărul de erori în cazul unui test video*

	Sleep deprivation		
	4 hours	12 hours	24 hours
Alcohol	16	18	22
	12	16	24
	17	25	32
No alcohol	11	13	12
	9	8	14
	12	11	12

22.1. Metoda two-way ANOVA pentru scoruri nerelaționate

Pasul 1

Introduceți datele. Cele două coduri pentru „Alcohol" (1 = Alcool și 2 = Fără alcool) se regăsesc în prima coloană.

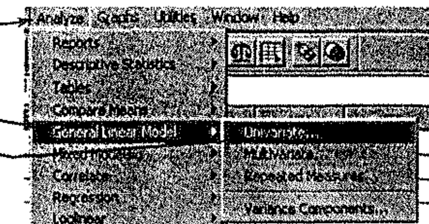
Cele trei coduri pentru „SleepDep" se regăsesc în a doua coloană (1 = 4 ore, 2 = 12 ore și 3 = 24 ore).

Greșelile sunt prezentate în a treia coloană. Etichetați aceste coduri după cum s-a precizat mai sus. Eliminați zecimalele din meniul „Variable View".

	Alcohol	SleepDep	Errors
1	1	1	16
2	1	1	12
3	1	1	17
4	1	2	18
5	1	2	16
6	1	2	25
7	1	3	22
8	1	3	24
9	1	3	32
10	2	1	11
11	2	1	9
12	2	1	12
13	2	2	13
14	2	2	8
15	2	2	11
16	2	3	12
17	2	3	14
18	2	3	12

Pasul 2

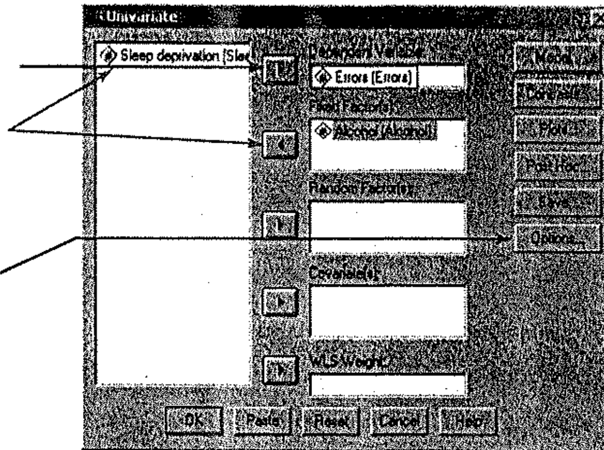
Selectați „Analyze", „General Linear Model" și „Univariate...".



Pasul 3

Selectați „Errors" și apăsați butonul ▶ de lângă caseta „Dependent Variable:" pentru a introduce parametrul acolo. Selectați „Alcohol" și „SleepDep" fie separat, fie împreună, și apăsați pe butonul „Fixed Factor(s):" pentru a le introduce în casetă.

Selectați „Options...".

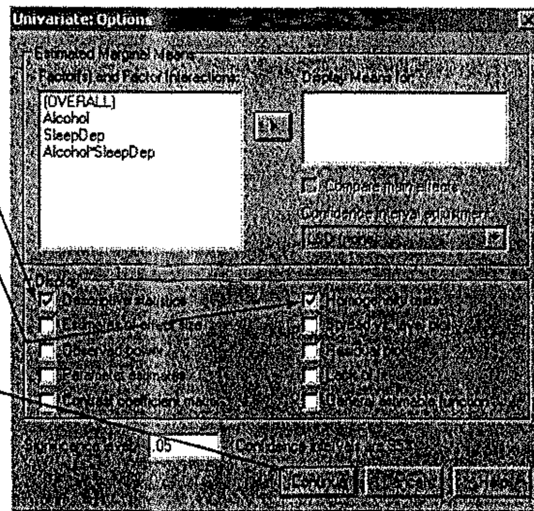


Pasul 4

Selectați „Descriptive statistics" și „Homogeneity tests".

Selectați „Continue".

În ecranul anterior, care va reapărea, selectați „Plots...".



Pasul 5

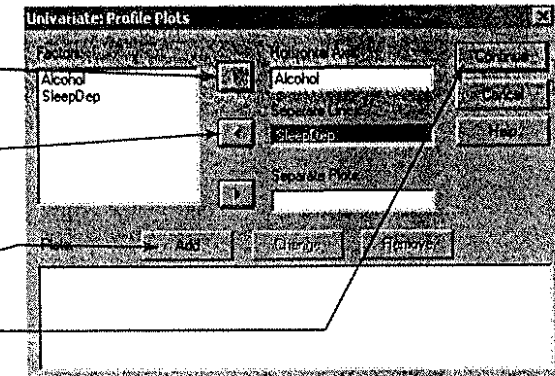
Selectați „Alcohol" și apăsați butonul ▶ de lângă caseta „Horizontal Axis:" pentru a introduce parametrul acolo.

Selectați „SleepDep" și apăsați butonul ▶ de lângă caseta „Separate Lines:" pentru a introduce opțiunea în casetă.

Selectați „Add".

Selectați „Continue".

Apăsați „OK" în ecranul anterior, care va reapărea.



22.2. Interpretarea output-ului

Al doilea tabel conține mediile, abaterile standard și numărul (N) de cazuri pentru cele două variabile „Alcohol" și „Sleep deprivation" luate separat sau împreună. Așadar, media pentru condiția „Alcohol" este comparată cu „Totalul" pentru „Sleep deprivation" (adică 20,22). Media pentru privarea de 4 ore de somn este comparată cu „Total" pentru „Alcohol" (adică 12,83).

Descriptive Statistics

Dependent Variable: Errors

Alcohol	Sleep deprivation	Mean	Std. Deviation	N
Alcohol	4 hrs	15.00	2.646	3
	12 hrs	19.67	4.726	3
	24 hrs	26.00	5.292	3
	Total	20.22	6.099	9
No alcohol	4 hrs	10.67	1.528	3
	12 hrs	10.67	2.517	3
	24 hrs	12.67	1.155	3
	Total	11.33	1.871	9
Total	4 hrs	12.83	3.061	6
	12 hrs	15.17	5.981	6
	24 hrs	19.33	8.066	6
	Total	15.78	6.330	18

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Errors

F	df1	df2	Sig.
2.786	5	12	.068

Testează ipoteza de nul conform căreia varianța de eroare pentru variabila dependentă este aceeași pentru toate grupele.

a. Model: Intercept + Alcohol + SleepDep + Alcohol*SleepDep

Al treilea tabel ne oferă informații referitoare la testul Levene, care verifică similaritatea varianțelor. Din moment ce semnificația acestui test este de 0,068 (valoare care este mai mare de 0,05), varianțele sunt similare. Dacă testul ar fi fost semnificativ, valorile ar trebui să fie transformate, de pildă, prin folosirea unei scale logaritmice pentru producerea similarității varianțelor. Aceasta este o chestiune de încercare și eroare – încercați diferite transformări până când varianțele devin similare.

Tests of Between-Subjects Effects

Dependent Variable: Errors

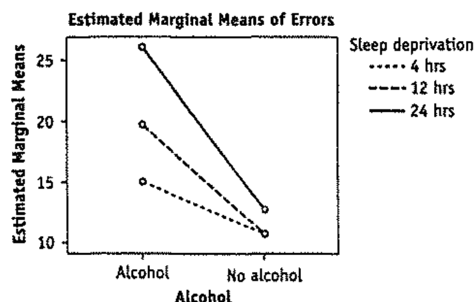
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	546.444a	5	109.289	9.739	.001
Intercept	4480.889	1	4480.889	399.267	.000
Alcohol	355.556	1	355.556	31.683	.000
SleepDep	130.111	2	65.056	5.797	.017
Alcohol * SleepDep	60.778	2	30.389	2.708	.107
Error	134.667	12	11.222		
Total	5162.000	18			
Corrected Total	681.111	17			

a. Coeficientul R^2 = 802 (R^2 ajustat = 0,720)

Al patrulea tabel indică nivelele de semnificaţie pentru cele două variabile, „Alcohol" şi „SleepDep", şi interacţiunea dintre acestea.

- În tabelul de analiză a varianţei, Raportul *F* pentru cele două efecte principale („Alcohol" şi „SleepDep") este prezentat primul.
- Pentru prima variabilă, cea a alcoolului, Raportul *F* este 31,683, ceea ce este semnificativ la nivelul 0,0005. Deoarece sunt numai două condiţii pentru ca acest efect să se producă, putem conchide că scorul mediu al uneia dintre condiţii este mult mai mare decât pentru cealaltă
- Pentru a doua variabilă a deprivării de somn, aceasta este egală cu 5,797, valoare care are un nivel exact de semnificaţie de 0,017. Cu alte cuvinte, Raportul *F* este semnificativ din punct de vedere statistic la nivel de 0,05, ceea ce înseamnă că mediile celor trei condiţii legate de somn nu sunt similare.
- Mediile care diferă de celelalte pot fi determinate ulterior prin folosirea testelor prin comparaţii multiple, cum este Testul *t* pentru eşantioane independente.
- Raportul *F* pentru interacţiunile dintre cele două variabile („Alcohol*SleepDep") este de 2,708. Din moment ce nivelul exact de semnificaţie al acestui raport este de 0,107, tragem concluzia că nu există nici o interacţiune semnificativă.

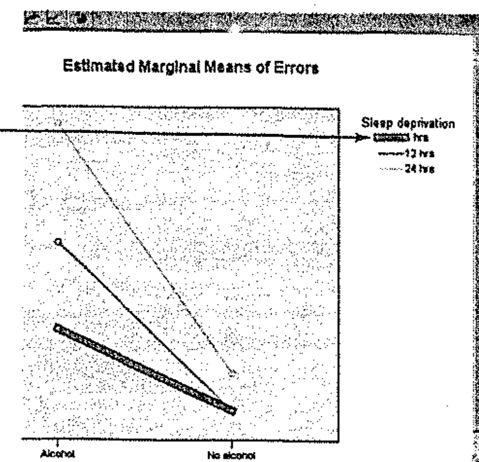
Acest grafic este realizat pentru mediile celor şase condiţii. A fost editat cu ajutorul comenzii „Chart Editor". Stilul liniilor colorate diferit a fost schimbat astfel încât să fie mai uşor de deosebit.



Pasul 1

Pentru a modifica stilul liniei, daţi dublu clic pe plot, pentru a selecta meniul „Chart Editor".

Selectaţi linia din legendă pe care doriţi să o schimbaţi.

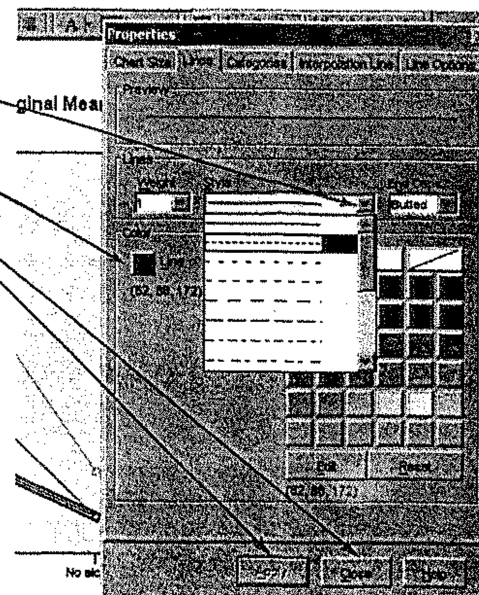


Pasul 2

Apăsaţi butonul ▾ de lângă „Style" şi alegeţi stilul pe care-l doriţi pentru linia respectivă. Modificaţi culoarea dacă doriţi.

Selectaţi „Apply" şi apoi „Close".

Selectaţi următoarea linie pe care doriţi să o editaţi şi repetaţi aceiaşi paşi.



22.3. Raportarea output-ului

- Putem raporta rezultatele din output după cum urmează: „O analiză de varianţă two-way ANOVA demonstrează obţinerea unor efecte semnificative în cazul alcoolului ($F_{2,12}$ = 31,68, p < 0,001) şi al privării de somn ($F_{2,12}$ = 5,80, p = 0,017), dar nu şi în cazul interacţiunii dintre cele două variabile ($F_{2,12}$ = 2,70, p = 0,107)".
- De obicei, se realizează şi un tabel sumar al analizei de varianţă. Unul simplu, asemănător celui din tabelul 22.2, ar elimina unele informaţii care nu sunt necesare din al treilea tabel din output.
- Din cauza faptului că factorul „SleepDep" are mai mult de două condiţii, trebuie să folosim un test prin comparaţii multiple adecvat, pentru a determina mediile ale căror grupuri diferă semnificativ (vezi capitolele 21 şi 23).
- De asemenea, trebuie să raportăm mediile şi abaterile standard ale grupurilor care diferă semnificativ. Aceste statistici descriptive sunt date în al doilea tabel al output-ului.

Tabelul 22.2. *Tabel sumar al analizei de varianţă*

Source of variation	Sums of squares	Degrees of freedom	Mean square	F-ratio	Probability
Alcohol	355.56	1	355.56	31.68	< 0.001
Sleep deprivation	130.11	2	65.06	5.80	< 0.05
Alcohol with sleep deprivation	60.78	2	30.39	2.71	not significant
Error	134.67	12	11.22		

23 Comparaţii multiple cu ANOVA

Rezumat

- Acest capitol extinde aria de acoperire a Testelor *t* multiple din capitolele 20 şi 21. Explică modul în care se decide care perechi de medii sunt diferite în mod semnificativ între ele, în cazul analizei de varianţă.
- Tehnica se foloseşte atunci când aveţi mai mult de două medii. Nu adaugă informaţii suplimentare dacă există doar două valori medii.
- Procedurile ANOVA cu mai mult de două perechi de medii vor avea cu siguranţă de profitat de pe urma folosirii metodelor de comparaţie multiplă.
- Este imposibil să se ofere o regulă general valabilă referitoare la testul prin comparaţii multiple care ar trebui folosit în diferite situaţii, deoarece literatura de specialitate nu s-a pus de acord în această privinţă.

Este important să ştiţi cu exactitate în ce constau diferenţele semnificative dintre diferitele condiţii ale studiului dumneavoastră. Curentul general în studiile ANOVA vă poate oferi o parte din adevăr. SPSS conţine un număr de proceduri ulterioare care se aplică, desigur, după colectarea datelor şi nu sunt planificate de la început. Toate acestea funcţionează în moduri uşor diferite. Există o discuţie detaliată pe această temă în Howell (2002). Vom ilustra folosirea acestor proceduri de comparaţii multiple (folosind informaţiile din tabelul 23.1, care au fost discutate anterior în capitolul 20).

Tabelul 23.1. *Informaţii pentru un studiu asupra efectelor hormonilor*

Group 1 Hormone 1	Group 2 Hormone 2	Group 3 Placebo control
9	4	3
12	2	6
8	5	3

23.1. Testele prin comparaţii multiple

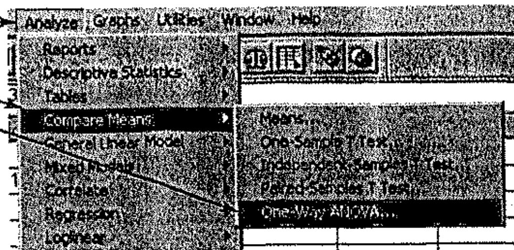
Pasul 1

Selectaţi fişierul cu date dacă îl aveţi salvat. Dacă nu, introduceţi datele.

	Condition	Depression
1	1	9
2	1	12
3	1	8
4	2	4
5	2	2
6	2	5
7	3	3
8	3	6
9	3	3

Pasul 2

Selectaţi opţiunile „Analize", „Compare Means" şi „One-way ANOVA".

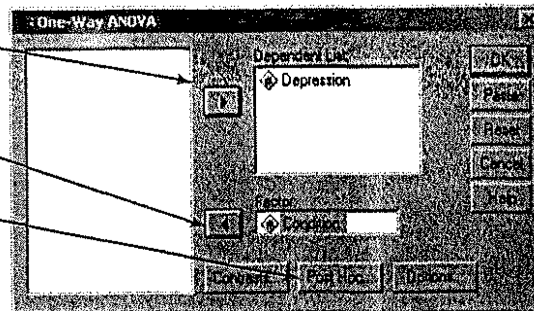


Pasul 3

Selectaţi opţiunea „Depression" şi apăsaţi butonul ► de lângă caseta „Dependent List:" pentru a introduce această opţiune în casetă.

Selectaţi opţiunea „Condition" şi apăsaţi butonul ◄ de lângă caseta „Factor:" pentru a introduce această opţiune în casetă.

Alegeţi opţiunea „Post Hoc...".

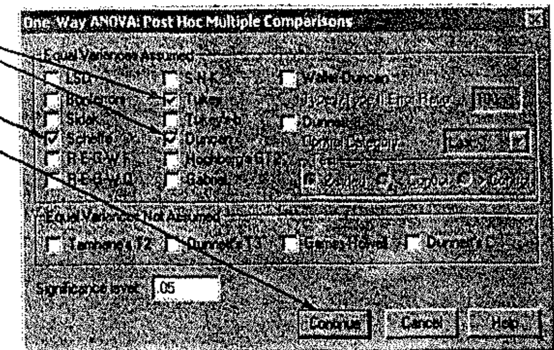


Pasul 4

Selectaţi „Tukey", „Duncan" şi „Scheffe".

Alegeţi „Continue".

Selectaţi „OK" în ecranul anterior, care va apărea din nou.



23.2. Interpretarea output-ului

ANOVA

Depression

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	68.222	2	34.111	10.586	.011
Within Groups	19.333	6	3.222		
Total	87.556	8			

Primul tabel reprezintă rezultatele analizei de varianţă. Raportul F pentru efectul dintre grupuri (adică efectul hormonilor) este 10,586, care are un nivel exact de semnificaţie de 0,011. Cu alte cuvinte, efectul dintre grupuri este semnificativ. Per total, mediile pentru cele trei grupuri diferă.

Tabelele al doilea şi ultimul oferă rezultatele celor trei teste de comparaţii multiple.

Multiple Comparisons

Dependent Variable: Depression

	(I) Condition	(J) Condition	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Tukey HSD	Hormone 1	Hormone 2	6.000*	1.466	.015	1.50	10.50
		Placebo control	5.667*	1.466	.019	1.17	10.16
	Hormone 2	Hormone 1	-6.000*	1.466	.015	-10.50	-1.50
		Placebo control	-.333	1.466	.972	-4.83	4.16
	Placebo control	Hormone 1	-5.667*	1.466	.019	-10.16	-1.17
		Hormone 2	.333	1.466	.972	-4.16	4.83
Scheffe	Hormone 1	Hormone 2	6.000*	1.466	.018	1.30	10.70
		Placebo control	5.667*	1.466	.023	.97	10.37
	Hormone 2	Hormone 1	-6.000*	1.466	.018	-10.70	-1.30
		Placebo control	-.333	1.466	.975	-5.03	4.37
	Placebo control	Hormone 1	-5.667*	1.466	.023	-10.37	-.97
		Hormone 2	.333	1.466	.975	-4.37	5.03

* Diferenţa dintre medii este semnificativă la nivelul de ,05

De exemplu, dacă folosim Testul Tukey al diferenţei semnificative oneste (HSD), media la grupul hormonului 1 este semnificativ diferită de media la grupul hormonului 2 (semnificaţia = 0,015) şi media grupului Controlului placebo (semnificaţia = 0,019).

Subgrupele omogene

Depression

	Condition	N	Subset for alpha = .05: 1	Subset for alpha = .05: 2
Tukey HSD[a]	Hormone 2	3	3.67	
	Placebo control	3	4.00	
	Hormone 1	3		9.67
	Sig.		.972	1.000
Duncan[a]	Hormone 2	3	3.67	
	Placebo control	3	4.00	
	Hormone 1	3		9.67
	Sig.		.828	1.000
Scheffe[a]	Hormone 2	3	3.67	
	Placebo control	3	4.00	
	Hormone 1	3		9.67
	Sig.		.975	1.000

Grupul hormonului 2 şi al controlului placebo aparţin aceleiaşi subgrupe, ceea ce înseamnă că nu sunt semnificativ diferite.

Grupul hormonului 1 este singurul din această a doua subgrupă. Aşadar, aceasta este semnificativ diferită de mediile celorlalte două grupe.

Sunt reprezentate mediile grupelor din subgrupele omogene.
a. Foloseşte Media armonică a dimensiunilor eşantioanelor = 3,000.

- Tabelul final, intitutlat „Subgrupele omogene", enumeră grupurile de medii care nu sunt semnificativ *diferite* între ele. Aşadar, dacă luăm aceste rânduri pentru a le folosi în Testul Tukey HSD, în acest caz există două subgrupe de medii. Subgrupa 1 indică faptul că mediile grupelor hormonului 2 şi ale controlului placebo cu valorile de 3,67 şi 4,00 nu diferă semnificativ. Subgrupa 2 conţine doar media grupei hormonului 1, în valoare de 9,67. Aşadar, media grupului hormonului 1 este semnificativ diferită atât de media grupei hormonului 2, cât şi de cea a grupelor de control placebo. Totuşi, aceasta înseamnă că grupele hormonului 2 şi cele ale controlului placebo nu sunt semnificativ diferite. În acest fel, modelul este identic şi în cazul Testelor Duncan şi Scheffé, deşi acest lucru nu este întotdeauna adevărat.
- Aşadar, toate cele trei teste prin comparaţii multiple sugerează acelaşi lucru : faptul că există diferenţe semnificative între grupele hormonului 1 şi hormonului 2 şi între grupele hormonului 1 şi cele ale controlului placebo. Alte diferenţe nu mai există. Aşadar, nu este posibil, de exemplu, să spunem că grupele hormonului 1 şi hormonului 2 sunt semnificativ diferite.
- Alegerea unuia dintre cele trei teste nu este una uşoară. Howell (2002) ne oferă câteva recomandări.

23.3. Raportarea output-ului

Putem raporta rezultatele output-ului după cum urmează : „O analiză de varianţă unifactorială pentru scoruri necorelate a demonstrat producerea unui efect general semnificativ pe tip de tratament medicamentos ($F_{2,6}$ = 10,59, p = 0,011). Prin Metoda Scheffé de analiză a intervalului s-a descoperit că grupul hormonului 1 era diferit de grupul hormonului 2 (p = 0,018) şi de grupul controlului placebo (p = 0,023), dar nu s-au mai găsit alte diferenţe semnificative".

Lucrări de referinţă

Howell, D. (2002), *Statistical Methods for Psychology* (ediţia a V-a, Duxbury Press : Boston).

24 Analiza de covarianță (ANCOVA) și analiza mixtă de varianță (ANOVA) 2-way

Rezumat

- Analiza de covarianță îi permite cercetătorului să controleze și să facă adaptări pentru variabilele care se corelează cu variabila dependentă pe care o aveți, prin compararea mediilor de la nivelul variabilei dependente. Aceste variabile sunt cunoscute ca fiind covariate ale variabilei dependente.
- În măsura în care nivelele variabilelor covariate sunt diferite în funcție de condițiile cercetării dumneavoastră, dacă nu vă ajustați variabila dependentă, variabilele covariate, veți încurca efectele variabilelor independente pe care le aveți cu influența exercitată de către diferențele preexistente dintre condițiile produse de diferite nivele ale covariatelor.
- Prin controlul variabilelor covariate, nu faceți decât să îndepărtați efectul pe care acestea îl au asupra scorurilor pe care le dețineți în ceea ce privește variabila dependentă. Așadar, după ce ați făcut ajustările necesare pentru variabilele covariate, variația care rămâne între condiții nu se poate datora covariatelor.
- Una dintre utilizările obișnuite ale metodei ANCOVA se regăsește la nivelul proiectelor anterioare și ulterioare testelor. Presupunând că demersul anterior testului vă sugerează că diferitele condiții ale experimentului au medii diferite înaintea testării (de exemplu, grupurile experimentale și cele de control sunt diferite), metoda ANCOVA poate fi folosită pentru ajustarea acestor diferențe.
- Un model de analiză mixtă a varianței nu este decât un model de cercetare bifactorială (sau trifactorială etc.) care conține atât variabile pentru eșantioane independente, cât și variabile pentru eșantioane perechi.
- Modelele mixte generează măsuri cu mult mai complexe de estimare a erorilor, în comparație cu alte forme ale metodei ANOVA. Aceasta înseamnă că trebuie să aveți mare grijă atunci când realizați tabelele de sumar corespunzătoare.

Analiza covarianței este în mare parte identică analizei de varianță tratată în alte lucrări, având însă o diferență esențială. Aceasta constă în faptul că efectele variabilelor adiționale (covariatelor) sunt îndepărtate și nu mai fac parte din analiză. Este ca și cum am folosi corelația parțială pentru a scăpa de efectele unei a treia variabile asupra unei corelații. Vom ilustra procesarea unei analize a covarianței (ANCOVA), folosind datele din tabelul 24.1, care sunt aceleași cu cele din tabelul 20.1, cu diferența că au fost incluse și scorurile depresiei, măsurate chiar înainte de cele trei tratamente.

Tabelul 24.1. *Date pentru studiul efectelor hormonilor (analiza covarianței)*

Group 1 Hormone 1		Group 2 Hormone 2		Group 3 Placebo control	
Pre	Post	Pre	Post	Pre	Post
5	9	3	4	2	3
4	12	2	2	3	6
6	8	1	5	2	3

Este posibil ca diferențele înregistrate în nivelul depresiei înainte de tratament să afecteze rezultatul final al analizei. Practic, prin ajustarea scorurilor variabilei dependente pentru eliminarea acestor diferențe preexistente, este posibil să nu mai luăm în considerare posibilitatea ca aceste diferențe preexistente să afecteze analiza. Așadar, dacă (a) scorurile anterioare tratamentului sau testului sunt corelate cu scorurile ulterioare tratamentului sau testului, iar (b) scorurile anterioare testului sunt diferite în cazul celor trei tratamente, atunci aceste diferențe sesizate înaintea testului pot fi controlate din punct de vedere statistic prin eliminarea lor din analiză, folosind covarianța.

24.1. Analiza one-way ANCOVA

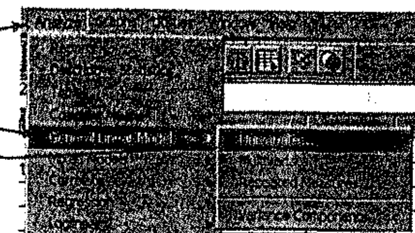
Pasul 1

Introduceți datele. Primele două coloane conțin date similare celor din capi-tolele 20 și 23. Datele din cea de-a treia coloană sunt noi.

	Condition	Posttest	Pretest
1	1	9	5
2	1	12	4
3	1	8	6
4	2	4	3
5	2	2	2
6	2	5	1
7	3	3	2
8	3	6	3
9	3	3	2

Pasul 2

Selectați opțiunile „Analyze", „General Linear Model" și „Univariate..."



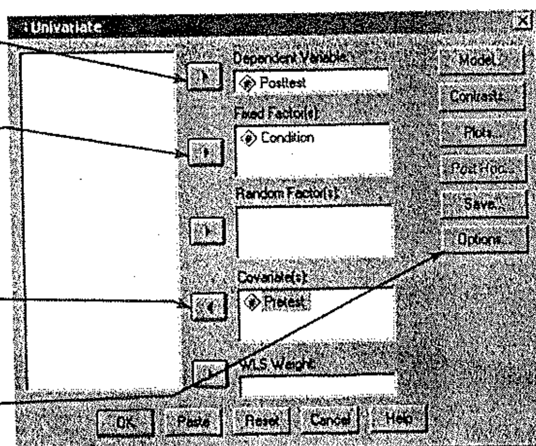
Pasul 3

Selectați opțiunea „Posttest" și apăsați butonul ▶ de lângă caseta „Dependent Variable:" pentru a introduce această opțiune în casetă.

Selectați opțiunea „Condition" și apăsați butonul ▶ de lângă caseta „Fixed Factor(s):" pentru a intro-duce această opțiune în casetă.

Selectați opțiunea „Pretest" și apăsați butonul ◀ de lângă caseta „Covariate(s):" pentru a introduce această opțiune în casetă.

Selectați „Options...".



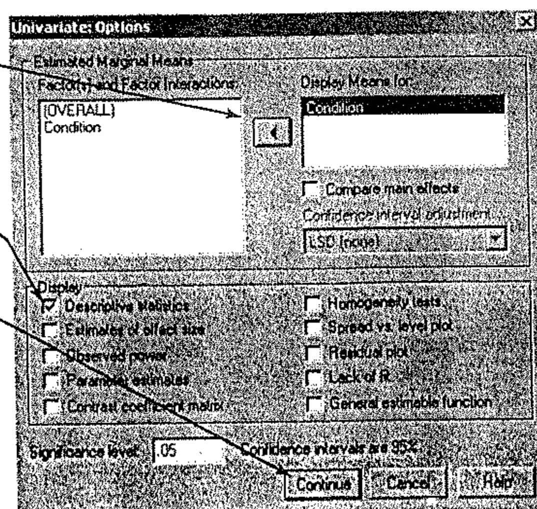
Pasul 4

Selectați „Condition" și apăsați butonul ◀ pentru a introduce opțiunea în caseta „Display Means for:".

Selectați „Descriptive statistics".

Selectați „Continue".

Apăsați „OK" în ecranul anterior, care reapare.



24.2. Interpretarea output-ului

Al doilea tabel conține mediile neajustate pentru cele trei situații.

Acest tabel oferă doar mediile pentru cele trei situații ulterioare testului. Totalul reprezintă media scorurilor ulterioare testului în condițiile datelor inițiale.

Descriptive Statistics

Dependent Variable: Posttest

Condition	Mean	Std. Deviation	N
Hormone 1	9.67	2.082	3
Hormone 2	3.67	1.528	3
Placebo control	4.00	1.732	3
Total	5.78	3.308	9

Al patrulea și cel din urmă tabel conține mediile ajustate în cele trei situații. Mediile ajustate în urma celor trei tratamente reprezintă valorile mediilor atunci când toate grupurile sunt ajustate astfel încât să fie identice la nivelul variabilelor covariate (în acest caz, la nivelul scorurilor depresiei dinaintea tratamentului).

Mediile ulterioare testului prezentate în acest tabel al output-ului au fost ajustate pentru a contracara efectul covariatelor asupra celor trei situații. Efectul variabilelor covariate a fost, practic, îndepărtat din aceste date.

Condition

Dependent Variable: Posttest

Condition	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Hormone 1	10.881ª	1.955	5.856	15.906
Hormone 2	2.952ª	1.443	-.756	6.661
Placebo control	3.500ª	1.269	.237	6.763

a. Convariatele care apar în model sunt evaluate la următoarele valori: Pretest = 3,11.

- Media ajustată este de 10,881 în cazul primului tratament, 2,952 în cazul celui de-al doilea și 3,500 în cazul celui de-al treilea tratament.
- Se observă că aceste medii ajustate par a fi diferite de cele neajustate incluse în al doilea tabel de output. În cazul primului tratament, media ajustată este de 10,88, iar cea neajustată este de 9,67. În cazul celui de-al doilea tratament, media ajustată este de 2,95, iar cea neajustată de 3,67, pe când în cazul celui de-al treilea tratament, media ajustată este de 3,50, în timp ce valoarea medie neajustată este de 4,00.

Al treilea tabel conține valorile Raportului F pentru analiza covarianței.

Tests of Between-Subjects Effects

Dependent Variable: Posttest

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	70.151[a]	3	23.384	6.718	.033
Intercept	27.684	1	27.684	7.953	.037
Pretest	1.929	1	1.929	.554	.490
Condition	26.425	2	13.213	3.796	.099
Error	17.405	5	3.481		
Total	388.000	9			
Corrected Total	87.556	8			

a. R Squared = .801 (Adjusted R Squared = .682)

În urma îndepărtării efectelor covariatei, nu mai există o diferență semnificativă între mediile din cele trei situații, deoarece semnificația este de 0,099, care nu este semnificativă din punct de vedere statistic.

- Raportul F al efectului principal are valoarea de 3,796 (13,213/3,481 = 3,796).
- Probabilitatea acestui Raport F este de 0,099. Cu alte cuvinte, este mai mare decât valoarea critică de 0,05 și, în consecință, nu este semnificativ din punct de vedere statistic.

24.3. Raportarea output-ului

- Putem raporta rezultatele output-ului după cum urmează: „Metoda one-way ANCOVA a demonstrat că, în momentul în care din evaluarea depresiei anterioare testului a fost eliminat efectul covariatei, efectul principal al tratamentului asupra depresiei ulterioare testului nu a fost semnificativ ($F_{2,5} = 3,79$, $p = 0,099$). Variabila covariată – scorurile depresiei anterioare tratamentului – avea un efect semnificativ asupra scorurilor depresiei ulterioare tratamentului". În mod normal, ar trebui să raportați și schimbările de la nivelul valorilor medii, realizate în momentul în care covariata a fost îndepărtată.
- În plus, în mod obișnuit se introduce și un tabel sumar ANCOVA, cum este tabelul 24.2.

Tabelul 24.2. *Tabelul sumar ANCOVA al efectelor tratamentelor pentru controlarea depresiei asupra depresiei anterioare tratamentelor*

Source of variance	Sums of squares	Degrees of freedom	Mean square	F-ratio
Covariate (pre-treatment depression scores)	43.73	1	43.73	12.56*
Main effect (treatment)	26.43	2	13.21	3.80
Residual error	17.41	5	3.48	

*Semnificativ la un nivel de 5%.

24.4. Modelul 2-way ANOVA mixt

O analiză two-way mixtă realizată asupra varianței are un factor care presupune eșantioane independente (scoruri necorelate) și unul care presupune eșantioane perechi (scoruri corelate). Factorii sunt variabile independente. Această analiză este ilustrată folosind datele din tabelul 24.3 (*ISP*, tabelul 24.6), care conține scorurile respectului de sine la copii, măsurate înainte și după o manipulare experimentală în cadrul căreia jumătate dintre copii (aleși în mod aleatoriu) au fost lăudați pentru comportamentul lor (condiție experimentală), iar celeilalte jumătăți nu i s-a oferit nici un fel de feedback (condiție de control).

Tabelul 24.3. *Nivelele respectului de sine anterioare și ulterioare testului, în două condiții*

Music score	Mathematics score
2	8
6	3
4	9
5	7
7	2
7	3
2	9
3	8
5	6
4	7

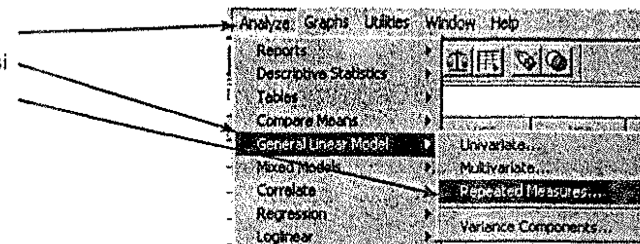
Pasul 1

Introduceți datele. Codurile pentru cele două condiții (1 = Condiție de control și 2 = Condiție experimentală) se regăsesc în prima coloană. Etichetați aceste valori ca atare. Valorile numerice anterioare și ulterioare testului se află în coloanele a doua și a treia.

	Condition	Pretest	Posttest
1	1	6	5
2	1	4	6
3	1	5	7
4	2	7	10
5	2	5	11
6	2	5	12
7			

Pasul 2

Selectați „Analyze", „General Linear Model" și „Repeated measures...".

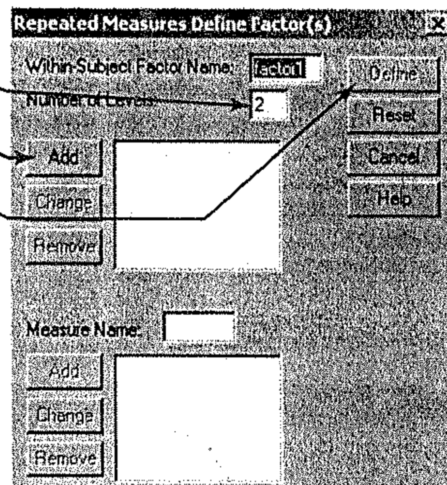


Pasul 3

Tastați „2" în caseta „Number of Levels:".

Selectați „Add".

Selectați „Define".

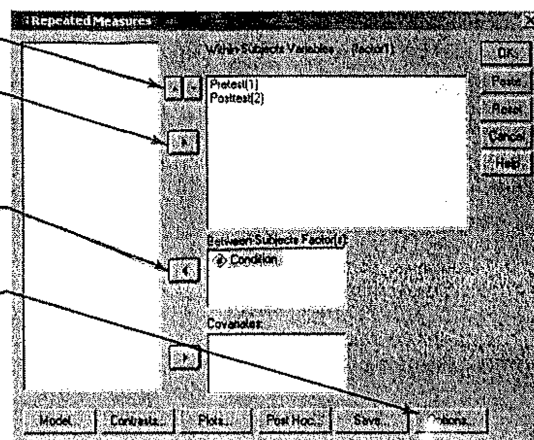


Pasul 4

Selectați „Pretest" și „Posttest" fie separat, fie împreună, apoi apăsați butonul ▶ de lângă caseta „Within-Subjects Variables:" pentru a introduce aceste opțiuni în casetă.

Selectați „Condition" și apăsați butonul ◀ de lângă caseta „Between-Subjects Factor(s):" pentru a introduce opțiunea în casetă.

Selectați „Options...".

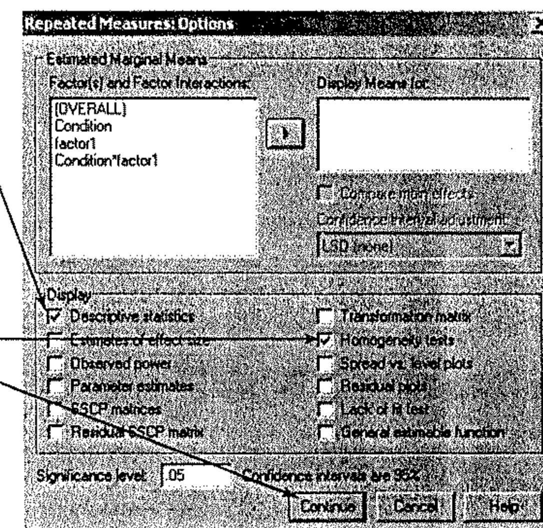


Pasul 5

Selectați „Descriptive statistics" și „Homogeneity tests".

Selectați „Continue".

În caseta anterioară, care va reapărea, selectați „Plots...".



Pasul 6

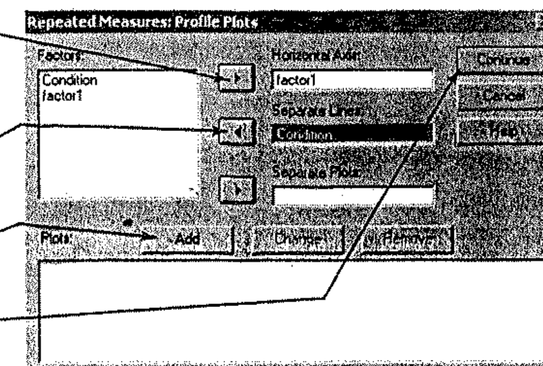
Selectați „factor1" („Pretest" „Posttest") și apăsați butonul ▶ de lângă caseta „Horizontal Axis:" pentru a adăuga opțiunea în casetă.

Selectați opțiunea „Condition" și apăsați butonul ◀ de lângă caseta „Separate Lines:" pentru a introduce opțiunea în casetă.

Selectați „Add".

Selectați „Continue".

Apăsați „OK" în ecranul anterior, care va reapărea.



24.5. Interpretarea output-ului

Output-ul oferă zece tabele şi un plot. Aici sunt prezentate doar tabelele cele mai importante.

Al doilea tabel conţine media şi abaterea standard pentru cele două grupuri.

Descriptive Statistics

	Condition	Mean	Std. Deviation	N
Pretest	Control	5.00	1.000	3
	Experimental	5.67	1.155	3
	Total	5.33	1.033	6
Posttest	Control	6.00	1.000	3
	Experimental	11.00	1.000	3
	Total	8.50	2.881	6

Al treilea tabel arată măsura în care matricele covarianţelor ulterioare testelor sunt egale în cele două situaţii. Această analiză a varianţei presupune că sunt egale. Din moment ce nivelul de semnificaţie de 0,951 este mai mare de 0,05, matricele sunt similare şi această presupunere se dovedeşte a fi adevărată.

Box's Test of Equality of Covariance Matrices[a]

Box's M	.757
F	.115
df1	3
df2	2880.000
Sig.	.951

Verificaţi ca această valoare să *nu* fie mai mică de 0,05. Dacă valoarea nu este semnificativă (aşa cum este în acest caz), covarianţele grupurilor sunt egale în urma testelor.

Testează ipoteza de nul conform căreia matricele de covarianţă observate ale variabilelor dependente sunt egale de la un grup la altul.
a. Model: Interceptare + Condiţie
Model la nivelul subiecţilor: factor 1

Cel de-al nouălea tabel arată măsura în care varianţa de eroare a celor două variabile este similară în cele două condiţii. Un nivel de semnificaţie mai mare de 0,05 indică faptul că aceste varianţe sunt similare.

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
Pretest	.308	1	4	.609
Posttest	.000	1	4	1.000

Verificaţi această coloană pentru a vedea nivelul semnificaţiei. Semnificaţia înseamnă că varianţele de eroare sunt seminificativ diferite în cazul celor două sau mai multe condiţii – fie în condiţiile anterioare testului, fie ulterior acestuia.

Testează ipoteza de nul conform căreia varianţa de eroare a variabilelor dependente este egală de la un grup la altul.
a. Model: Interceptare + Condiţie
Model la nivelul subiecţilor: factor 1

Cel de-al optulea tabel conţine informaţii pentru Testul *F*. Testul *F* care ne interesează în mod deosebit este cel referitor la interacţiunea dintre factorii de la nivelul subiecţilor şi cei care acţionează între subiecţi (factor 1* Condiţie). Acest Raport *F* este de 7,682 şi are o valoare de probabilitate de 0,05. Cu alte cuvinte, această interacţiune este semnificativă la limită. Dacă ne uităm la mediile pentru cele patru grupuri, vom vedea că pe măsură ce media pentru condiţia de control înregistrează o creştere scăzută de la valoarea anterioară testului (5,00) la cea ulterioară testului (6,00), media pentru condiţia experimentală manifestă o creştere mai accentuată de la nivelul anterioară testului (5,67) la cel superioară testului (11,00).

Tests of Within-Subjects Contrasts

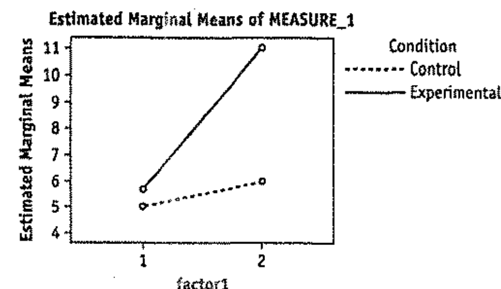
Measure: MEASURE_1

Source	factor1	Type III Sum of Squares	df	Mean Square	F	Sig.
factor1	Linear	30.083	1	30.083	16.409	.015
factor1 * Condition	Linear	14.083	1	14.083	7.682	.050
Error(factor1)	Linear	7.333	4	1.833		

Această valoare indică o interacţiune semnificativă la limită între valorile anterioară şi ulterioare testului pe de o parte şi condiţia experimentală pe de altă parte. Cu alte cuvinte, există între celule diferenţe care nu pot fi explicate de diferenţele preexistente dintre grupurile de participanţi sau pur şi simplu de schimbările tuturor condiţiilor între momentul anterior şi cel ulterior testului. Uitaţi-vă la tabelul 24.5 şi veţi avea o imagine mai clară asupra tendinţelor de la nivelul mediilor.

- Pentru a determina dacă aceste creşteri sunt semnificative din punct de vedere statistic, am putea să efectuăm un Test *t* pentru eşantioane perechi între scorurile anterioare şi ulterioare testului, separat pentru cele două condiţii (folosind corecţia Bonferroni pentru numărul de comparaţii pe care le vom efectua).
- De asemenea, putem vedea dacă cele două condiţii erau diferite la testul anterior şi la cel ulterior cu ajutorul unui Test *t* pentru eşantioane independente, efectuat separat pentru cele două perioade de testare.

Graficul prezentat în continuare ilustrează media celor patru celule, care vă poate ajuta să înţelegeţi mai uşor relaţia dintre ele. Acesta a fost editat pentru schimbarea stilului liniilor, aşa cum s-a explicat în capitolul 22.



24.6. Raportarea output-ului

- Putem raporta rezultatele output-ului după cum urmează: „Interacțiunea dintre cele două condiții și schimbările care au avut loc în timp au fost semnificative din punct de vedere statistic ($F_{1,4}$ = 7,68, p = 0,05). Deși mediile testului anterior nu erau semnificativ diferite (t = 0,76, DF = 4, two-tailed p = 0,492), media ulterioară testului în cazul condiției experimentale (M = 11,00, SD = 1,00) era cu mult mai ridicată (t = 6,12, DF = 4, two-tailed p = 0,004) decât cea înregistrată în cazul condiției de control (M=6,00, SD=1,00). Creșterea înregistrată de la testul anterior (M = 5,67, SD = 1,15) până la testul ulterior (M = 11,00, SC = 1,00) a fost semnificativă în cazul condiției experimentale (t = 4,44, DF = 2, two-tailed p = 0,047), dar nu și în cazul celei de control (t = 1,00, DF=2, two-tailed p = 0,423)".
- Tabelul 24.4 este un tabel de analiză a varianței pentru acest studiu.
- De asemenea, va fi utilă includerea unui tabel al mediilor (M) și al abaterilor standard (SD), așa cum se arată în tabelul 24.5.

Tabelul 24.5. *Tabelul sumar ANCOVA pentru un model bifactorial mixt*

Source of variance	Sums of squares	Degrees of freedom	Mean square	*F*-ratio
Between-subjects factor	24.08	1	24.08	72.25*
Between-subjects error	1.33	1	1.33	
Within-subjects factor	30.08	1	30.08	16.41*
Within-subjects error	7.33	4	1.83	
Interaction	14.08	1	14.08	7.68*

*Semnificativ la nivelul de 0,05.

Tabelul 24.5. *Mediile și abaterile standard pentru momentele anterioare și ulterioare testului, în cazul condițiilor de control și ale celor experimentale*

	Pre-test		Post-test	
Conditions	*M*	SD	*M*	SD
Control	5.00	1.00	6.00	1.00
Experimental	5.67	1.15	11.00	1.00

25 Citirea fișierelor ASCII sau a fișierelor text cu ajutorul opțiunii „Data Editor"

Rezumat

- Uneori, dețineți un fișier cu date pe care doriți să le folosiți în SPSS. Acest capitol vă explică modul în care puteți folosi date care încă nu sunt introduse în foaia de lucru „Data Editor" a SPSS-ului.
- Deși ca studenți veți avea rareori nevoie să folosiți fișiere ASCII, există un număr de baze de date cu informații arhivate pe care cercetătorii ar dori, probabil, să le analizeze.
- Merită să remarcăm faptul că uneori informațiile din alte foi de lucru pot fi introduse în SPSS. De exemplu, datele din foile de lucru Excel pot fi pur și simplu copiate și lipite în SPSS, dar acest lucru nu este posibil totuși cu întreaga foaie de lucru Excel. De asemenea, informațiile din SPSS pot fi salvate sub forma unui fișier Excel.

25.1. Introducere

SPSS pentru Windows este unul dintre numeroasele programe diferite pentru analizarea datelor. Există situații în care cercetătorii pot avea nevoie să folosească pachete informa-ționale pe care le-au pregătit pentru un alt computer și să ruleze acele informații în SPSS pentru Windows. Poate fi costisitor atât din punct de vedere al timpului, cât și din cel al banilor să reintroducem, de exemplu, în foaia de lucru „Data Editor" datele dintr-un sondaj de amploare. Uneori, singura formă în care sunt disponibile datele este aceea a unui fișier dintr-o arhivă electronică; este posibil să nu mai existe chestionarele inițiale. Indiferent de motivul pentru care se folosește un fișier cu date importate, SPSS pentru Windows poate accepta fișierele și sub alte forme. Mai precis, fișierele cu date sunt uneori scrise ca simplu text sau ca fișiere ASCII, din moment ce acestea pot fi transferate cu ușurință de la un tip de computer la altul. ASCII înseamnă „American Standard Code for Information Interchange". Pentru a analiza un fișier de date ASCII, trebuie mai întâi să îl citiți cu ajutorul opțiunii „Data Editor".

Să presupunem, de exemplu, că aveți un fișier de date ASCII numit „data.txt" care conține următoarele numere:

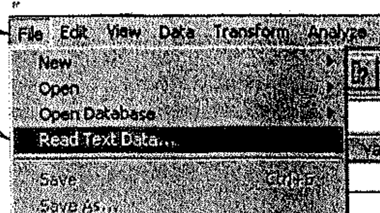
1118
2119
3218

Evident, nu puteţi folosi raţional un fişier ASCII înainte de a şti exact unde este informaţia pentru fiecare variabilă. Totuşi, noi ştim ce informaţii conţine micul fişier de mai sus şi unde se găsesc acestea. Cifrele din prima coloană nu fac decât să numeroteze trei participanţi diferiţi despre care deţinem informaţii. Valorile din cea de-a doua coloană conţin codul pentru gen, „1" reprezentând femeile şi „2" – bărbaţii, iar valorile din coloanele a treia şi a patra indică vârstele celor trei persoane. Vom efectua următoarea procedură pentru a introduce fişierul de date ASCII în „Data Editor".

25.2. Cum să citim un fişier de date ASCII sau un fişier text

Pasul 1

Selectaţi „File" şi „Read Text Data...".

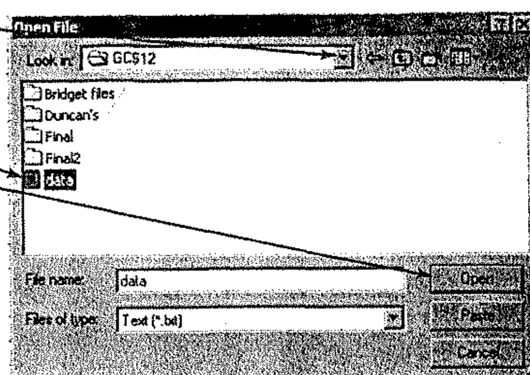


Pasul 2

Selectează butonul ▼ din caseta „Look in:" pentru a găsi fişierul de date (de exemplu, data).

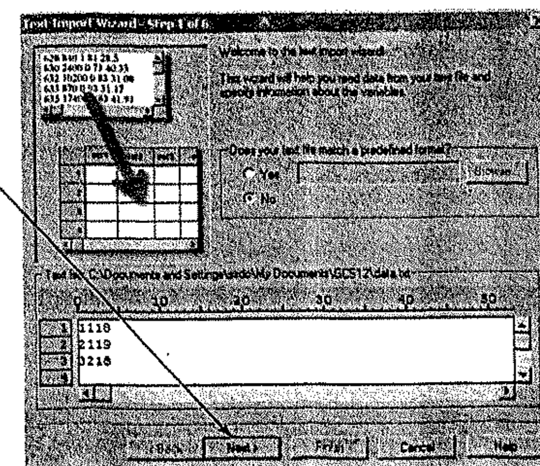
Selectaţi fişierul.

Selectaţi „Open".



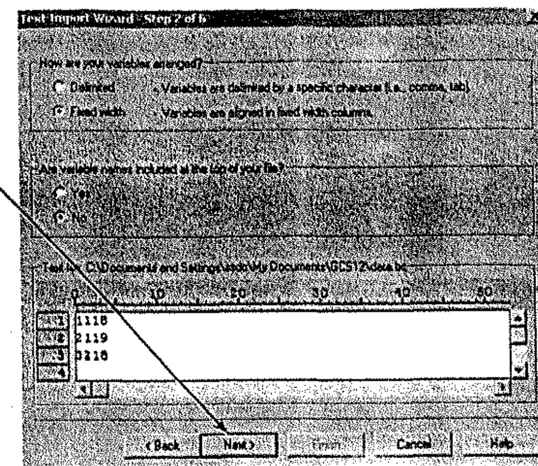
Pasul 3

Selectaţi „Next".



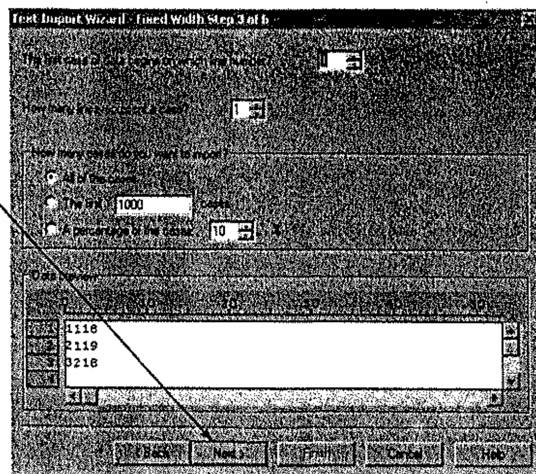
Pasul 4

Selectaţi „Next".



Pasul 5

Selectați „Next".



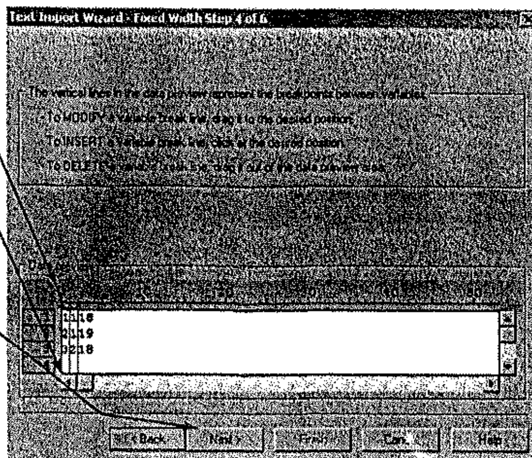
Pasul 6

Dați clic imediat după prima coloană, deoarece această coloană conține numărul participantului.

Dați clic după a doua coloană care conține codul genului.

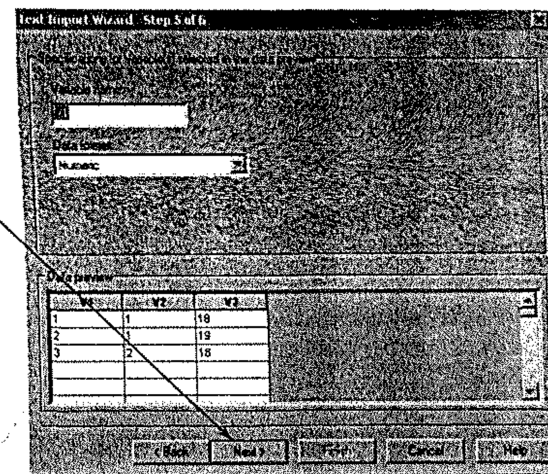
Ultimele două coloane conțin vârsta.

Selectați „Next".



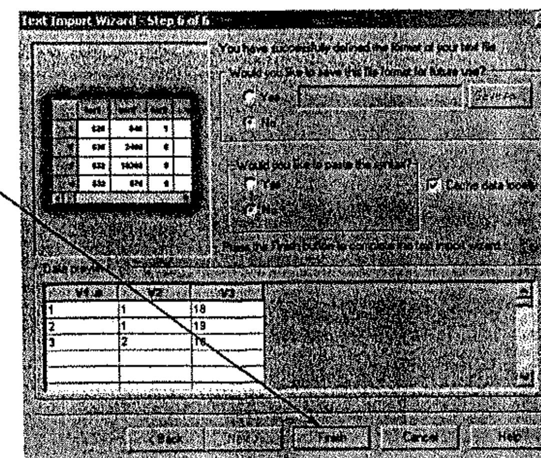
Pasul 7

Selectați „Next".



Pasul 8

Selectați „Finish".



Datele textului au fost introduse în „Data Editor" şi pot fi salvate sub formă de fişier SPSS.

	V1	V2	V3
1	1	1	18.0
2	2	1	19.0
3	3	2	18.0

26 Corelaţia parţială

Rezumat

- Dacă bănuiţi că o corelaţie dintre două variabile este afectată de corelaţia acestora cu o a treia variabilă, este posibil să faceţi adaptările necesare pentru efectele acestei variabile suplimentare prin folosirea procedurii de corelaţie parţială.
- Corelaţia dintre două variabile (înainte de parţializare) este cunoscută sub numele de corelaţie de ordin zero.
- Prin folosirea aplicaţiei SPSS, este posibilă şi controlarea simultană a mai multor variabile care pot afecta coeficientul de corelaţie.
- Dacă variabila pe care trebuie să o controlaţi este formată dintr-un număr mic de categorii nominale, este util să observaţi corelaţiile pe grupuri separate de cazuri, pe baza variabilei de control. De exemplu, dacă trebuie să controlaţi genul, atunci va trebui să vă separaţi eşantionul într-un subeşantion de bărbaţi şi într-unul de femei. Observaţi care sunt corelaţiile dintre variabilele principale în cazul acestor două grupuri. De multe ori, acest procedeu clarifică efectul parţializării într-un mod surprinzător.

Aplicaţia SPSS pentru Windows nu procesează cu uşurinţă corelaţiile parţiale dintr-o matrice de corelaţii de ordin zero (adică, dacă valorile corelaţiilor dintre variabilele analizate sunt introduse direct). În consecinţă, vom ilustra procesarea corelaţiilor parţiale cu scorurile brute din tabelul 26.1, care reprezintă scorurile la testul de inteligenţă numerică, la testul de inteligenţă verbală şi vârsta în ani. Vom corela scorurile la cele două teste, parţializând vârsta.

Tabelul 26.1. *Rezultatele la testele de inteligenţă numerică şi verbală şi vârsta*

Numerical scores	Verbal scores	Age
90	90	13
100	95	15
95	95	15
105	105	16
100	100	17

26.1. Corelaţia parţială

Pasul 1

În fereastra „Variable View" din „Data Editor", denumiţi cele trei coloane „Num_IQ", „Verb_IQ" şi „Age".

Înlăturaţi cele două zecimale prin introducerea valorii 0.

	Name	Type	Width	Decimals
1	Num_IQ	Numeric	8	0
2	Verb_IQ	Numeric	8	0
3	Age	Numeric	8	0

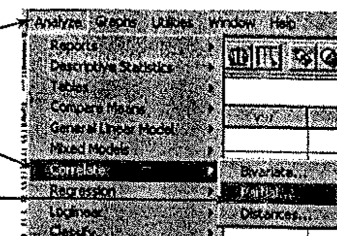
Pasul 2

În fereastra „Data View" din „Data Editor", introduceţi valorile IQ-ului numeric în prima coloană, cele ale IQ-ului verbal în a doua coloană, iar vârsta în cea de-a treia coloană.

	Num_IQ	Verb_IQ	Age
1	90	90	13
2	100	95	15
3	95	95	15
4	105	105	16
5	100	100	17

Pasul 3

Selectaţi comenzile „Analyze", „Correlate" şi „Partial...".

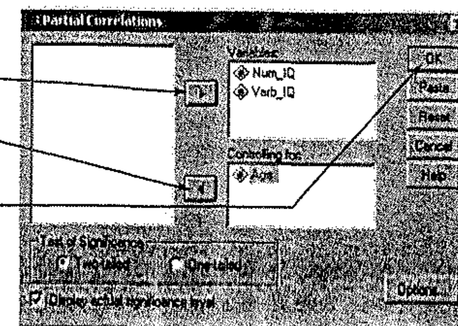


Pasul 4

Selectaţi „Num_IQ" şi „Verb_IQ", apoi apăsaţi butonul ▶ pentru a introduce aceste două variabile în caseta „Variables:".

Selectaţi „Age" şi apăsaţi butonul ▶ pentru a introduce variabila în caseta „Controlling for:".

Selectaţi „OK".



26.2. Interpretarea output-ului

Corelaţia parţială dintre valorile IQ-ului numeric şi ale celui verbal care ajută la controlarea variabilei vârstei este de 0,776.

Această celulă oferă aceleaşi informaţii ca şi celula din dreapta sus.

Correlations

Control Variables			Num_IQ	Verb_IQ
Age	Num_IQ	Correlation	1.000	.776
		Significance (2-tailed)	.	.224
		df	0	2
	Verb_IQ	Correlation	.776	1.000
		Significance (2-tailed)	.224	.
		df	2	0

Semnificaţia bilaterală este de 0,224, valoare care nu este semnificativă. Există două trepte de libertate.

26.3. Raportarea rezultatelor

Trebuie să menţionăm corelaţia iniţială dintre inteligenţa numerică şi cea verbală, care este de 0,92. Aşadar, am putea raporta rezultatele după cum urmează: „Corelaţia dintre inteligenţa numerică şi cea verbală este de 0,92 (*DF* = 3, two-tailed *p* = 0,025). Totuşi, corelaţia dintre inteligenţa numerică şi cea verbală atunci când este controlat parametrul vârstei scade la 0,78, valoare care nu este semnificativă (two-tailed *p* = 0,224). Cu alte cuvinte, nu există o relaţie semnificativă între inteligenţa numerică şi cea verbală atunci când parametrul vârstei este controlat".

27 Analiza factorială

Rezumat

- Există două tipuri de analiză factorială: analiza factorială exploratorie şi cea confirmatoare. Aplicaţia SPSS nu procesează analiza factorială confirmatoare în mod direct. Analiza factorială exploratorie este, probabil, cea mai importantă şi SPSS-ul deţineţi un pachet complet de opţiuni pentru aceasta.
- Analiza factorială (exploratorie) permite extragerea unor concluzii logice atunci când operăm cu un set complex de variabile, reducându-le la un număr mai mic de factori (sau supervariabile) care înlocuiesc multe dintre variabilele iniţiale. Deşi este posibil să obţinem informaţii valoroase dintr-o matrice de corelaţii dintre diferite variabile, dimensiunea mare a matricei poate îngreuna această operaţiune chiar şi atunci când operăm cu un număr relativ mic de variabile.
- Analiza factorială este folosită în mod obişnuit atunci când cercetătorul încearcă să înţeleagă modelul reacţiilor persoanelor care completează chestionare cu întrebări închise. Elementele care măsoară aspecte similare pot fi identificate prin analiză factorială şi, în consecinţă, pot fi deduse din structura răspunsurilor date la întrebările din chestionar.
- Totuşi, analiza factorială presupune o varietate de tehnici şi abordări pe care le puteţi considera prea complicate. Vă vom oferi o abordare „standard" care este folositoare în urmărirea obiectivelor majorităţii cercetătorilor.
- Analiza factorială necesită un oarecare discernământ, mai ales în ceea ce priveşte numărul de factori care trebuie extraşi. Viteza analizelor factoriale efectuate cu ajutorul computerului vă permite să încercaţi mai multe abordări, chiar în cadrul unei sesiuni scurte de analiză. Este bine să vedeţi care sunt rezultatele varierii metodei de analiză pentru a aprecia efectele pe care le are acest procedeu asupra concluziilor pe care le trageţi.

Procesarea unei analize factoriale a componentelor principale este exemplificată cu datele din tabelul 27.1, care reprezintă scorurile a şase variabile măsurate în cazul a nouă indivizi. Acest studiu are doar un rol ilustrativ; ar fi caraghios să efectuăm o analiză factorială pentru un număr atât de mic de cazuri. În mod obişnuit, ar trebui să aveţi în vedere lucrul cu un număr de cazuri care să reprezinte dublul sau triplul numărului de variabile de care dispuneţi. Următorul studiu este o analiză factorială standard care e potrivită majorităţii situaţiilor. Totuşi, aplicaţia SPSS are multe opţiuni pentru analiza factorială.

Tabelul 27.1. *Scorurile a șase variabile obținute de nouă persoane*

Individual	Batting	Crosswords	Darts	Scrabble	Juggling	Spelling
1	10	15	8	26	15	8
2	6	16	5	25	12	9
3	2	11	1	22	7	6
4	5	16	3	28	11	9
5	7	15	4	24	12	7
6	8	13	4	23	14	6
7	6	17	3	29	10	9
8	2	18	1	28	8	8
9	5	14	2	25	10	6

27.1. Analiza pe componente principale cu rotație ortogonală

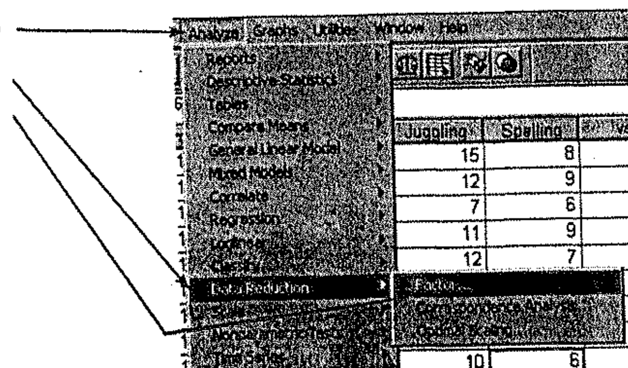
Pasul 1

Introduceți datele.

	Batting	Crosswords	Darts	Scrabble	Juggling	Spelling
1	10	15	8	26	15	8
2	6	16	5	25	12	9
3	2	11	1	22	7	6
4	5	16	3	28	11	9
5	7	15	4	24	12	7
6	8	13	4	23	14	6
7	6	17	3	29	10	9
8	2	18	1	28	8	8
9	5	14	2	25	10	6

Pasul 2

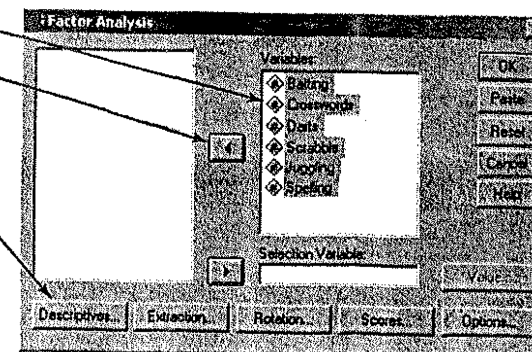
Selectați „Analyze", „Data Reduction" și „Factor...".



Pasul 3

Selectați cele șase variabile fie separat, fie împreună și apăsați butonul ◄ pentru a le introduce în caseta „Variables:".

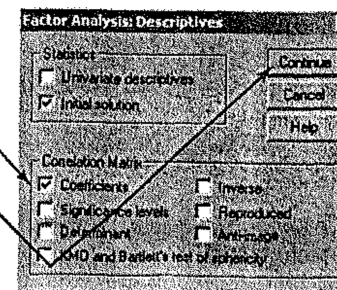
Selectați „Descriptives...".



Pasul 4

Selectați „Coefficients".

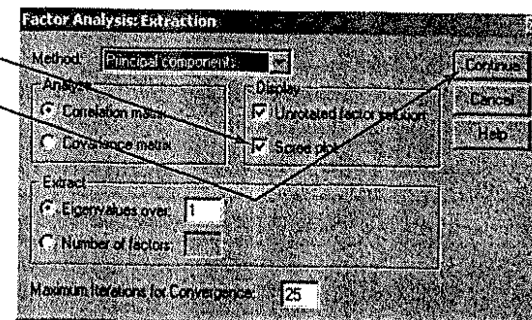
Selectați „Continue".



Pasul 5

Selectați „Scree plot".

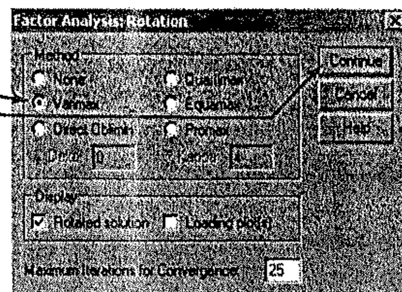
Selectați „Continue".



Pasul 6

Selectați „Varimax".

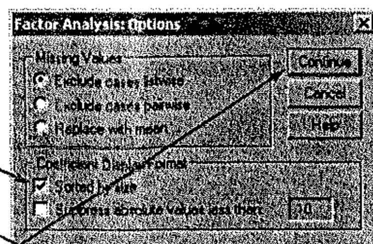
Selectați „Continue".



Pasul 7

Selectați „Sorted by size" și selectați „Continue".

Apăsați „OK" ca în ecranul inițial.



27.2. Interpretarea output-ului

- Primul tabel prezintă matricea de corelație. De aici, deducem că există două grupuri de variabile care sunt într-o relație puternică de corelație. Unul este format din jocul de baseball, jonglarea și jocul de darts, iar celălalt din cuvinte încrucișate, scrabble și jocuri de ortografiere a cuvintelor. Aceste grupuri au fost deja indicate, dar țineți minte că, din moment ce o matrice de corelație este simetrică, doar jumătatea inferioară situată sub diagonală a fost marcată. În mod obișnuit, în analiza factorială, matricea de corelație este mult mai greu de descifrat decât în acest caz. Datele cazului nostru au fost mult stilizate.

Correlation Matrix

		Batting	Crosswords	Darts	Scrabble	Juggling	Spelling
Correlation	Batting	1.000	.000	.910	-.047	.963	.096
	Crosswords	.000	1.000	.081	.883	.023	.795
	Darts	.910	.081	1.000	-.005	.902	.291
	Scrabble	-.047	.883	-.005	1.000	-.080	.789
	Juggling	.963	.023	.902	-.080	1.000	.108
	Spelling	.096	.795	.291	.789	.108	1.000

Deoarece analiza factorială presupune, de cele mai multe ori, o mulțime de variabile, iar suprafața ecranului computerului este limitată, de obicei este greu să privim matricea de corelație în întregime.

- Al treilea tabel ne arată că în acest caz au fost extrași inițial doi factori referitori la componentele principale. Computerul ignoră factorii eigenvalue mai mici decât 1,00. Acest lucru se întâmplă deoarece astfel de factori manifestă o variație de eroare care nu poate fi interpretată. Desigur, este posibil ca analiza dumneavoastră să aibă chiar mai mulți (sau mai puțini) factori.

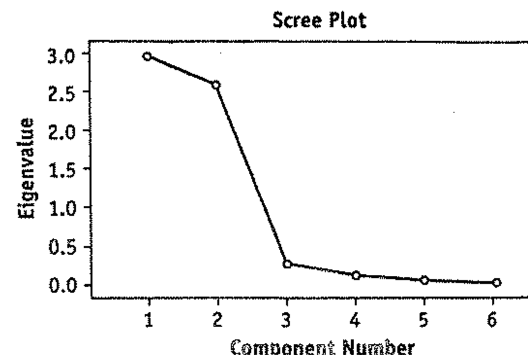
Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.951	49.186	49.186	2.951	49.186	49.186	2.876	47.931	47.931
2	2.579	42.981	92.167	2.579	42.981	92.167	2.654	44.236	92.167
3	.264	4.401	96.567						
4	.124	2.062	98.630						
5	.058	.974	99.604						
6	.024	.396	100.000						

Metodă de extragere: analiza componentelor principale.

Primii doi factori (componente) vor fi analizați mai departe de către computer deoarece au eigenvalues mai mare decât 1,00.

- Diagrama de tip scree plot ne arată că după al doilea factor apare o scădere a valorilor eigenvalue ale factorilor: linia este practic plată după al doilea factor. Din moment ce în analiza factorială este important să vă asigurați că nu aveți prea mulți factori, veți dori, probabil, să vă efectuați propria analiză factorială și rotația stipulând numărul de factori odată ce aveți rezultatele testului efectuat prin metoda grafică scree plot. (Acest lucru se poate realiza prin introducerea numărului respectiv în comanda „Number of factors : " din subcaseta de dialog „Factor Analysis : Extraction". În cazul datelor noastre, nu este nevoie să facem acest lucru, deoarece computerul a folosit primii doi factori și i-a ignorat pe ceilalți din cauza valorii eigenvalue minime necesare de 1,00. Nu este o practică rară să se proceseze din nou analiza pe componente, având în vedere modelul care rezultă.



- Aceste două componente sunt, mai apoi, rotite în plan ortogonal şi ponderile celor şase variabile în cei doi factori sunt ilustrate în cel de-al cincilea tabel intitulat „Rotated Component Matrix".

Rotated Component Matrix[a]

	Component	
	1	2
Batting	.980	−.012
Juggling	.979	−.011
Darts	.962	.104
Crosswords	.006	.951
Scrabble	−.078	.951
Spelling	.153	.914

Metoda de extragere: Analiza componentelor principale.
Metoda de rotire: Varimax cu normalizare Kaiser
a. Rotire în trei etape.

- Variabilele sunt ordonate şi sortate în funcţie de ponderea pe care o au în primul factor, de la cele care au cea mai mare pondere la cele cu ponderea cea mai redusă. Aceasta ajută la interpretarea factorului respectiv, deoarece elementele cu pondere mare sunt şi cele care vă ajută cel mai mult atunci când trebuie să identificaţi factorul respectiv.
- În primul factor, jocul de baseball are cea mai mare pondere (0,980), aceasta fiind urmată de cea exercitată de jonglerii (0,979) şi de jocul de darts (0,962).
- În cel de-al doilea factor, cea mai mare pondere o au cuvintele încrucişate (0,951), urmate de scrabble (0,951) şi de jocurile de ortografiere (0,914). Lipsa aparentă a diferenţei ponderii exercitate de cuvintele încrucişate şi de scrabble se datorează rotunjirii. Acest lucru se poate vedea dacă daţi dublu clic pe tabelul matricei componentelor rotite şi încă un dublu clic pe aceste două influenţe, la rândul lor.
- Vom interpreta semnificaţia acestor factori în funcţie de conţinutul variabilelor care le-au influenţat cel mai mult.
- Procentul varianţei pentru fiecare dintre factorii rotiţi ortogonal este dat în al treilea tabel în secţiunea „% of variance" în secţiunea „Rotation Sums of Squared Loadings". Această sumă este 47,931 pentru primul factor şi 44,236 pentru cel de-al doilea.

27.3. Raportarea output-ului

În mod obişnuit, factorii şi variabilele se tabulează dacă spaţiul disponibil o permite. Din moment ce informaţiile din exemplul nostru se referă la diferite teste de abilităţi, tabelul analizei factoriale ar putea fi similar tabelului 27.2. Cifrele au fost date cu o exactitate de două zecimale.

- Modalitatea exactă de raportare a rezultatelor unei analize factoriale depinde de scopul analizei. Una dintre modalităţile de descriere a rezultatelor ar fi următoarea: „S-a efectuat o analiză factorială pe componente principale pe baza corelaţiilor dintre

cele şase variabile. Iniţial, au fost extraşi doi factori care au coeficienţi eigenvalue mai mari sau egali cu 1,00. Rotirea ortogonală a factorilor a determinat structura factorială reprezentată în tabelul 27.2. Primul factor reprezenta 48% din varianţă, iar cel de-al doilea 44%. Primul factor pare a fi coordonarea ochi-mână, iar cel de-al doilea flexibilitatea verbală". Din cauza faptului că factorii trebuie să fie interpretaţi, pot apărea diferenţe de interpretare în analiza factorială.

Tabelul 27.2. *Matricea influenţei factorilor ortogonali pentru şase abilităţi*

Variable	Factor 1	Factor 2
Skill at batting	0.98	−0.01
Skill at crosswords	0.01	0.95
Skill at darts	0.96	0.10
Skill at Scrabble	−0.08	0.95
Skill at juggling	0.98	−0.01
Skill at spelling	0.15	0.91

28 Regresia multiplă stepwise

Rezumat

- Regresia multiplă stepwise este o modalitate de alegere a predictorilor unei anumite variabile dependente pe baza criteriilor statistice.
- Practic, procedura statistică hotărăşte care variabilă independentă reprezintă cel mai bun predictor, al doilea ca valoare etc.
- Se pune accentul pe identificarea celor mai buni predictori pentru fiecare etapă. Când predictorii se află într-o relaţie strânsă de corelaţie atât între ei, cât şi cu variabila dependentă, adeseori, o variabilă ajunge să fie enumerată ca fiind predictor, iar cealaltă nu. Acest lucru nu înseamnă că a doua variabilă *nu* constituie un predictor, ci doar că nu adaugă nici o informaţie în plus predicţiei faţă de cele oferite de primul predictor. Uneori, cel mai bun predictor este doar cu puţin mai bun decât al doilea în ordinea valorii şi unele variaţii minore la nivelul procedurilor pot afecta alegerea unuia dintre cei doi drept predictor.
- Există o serie de variante de regresie multiplă. Metoda stepwise este, de obicei, o alegere bună, deşi se pot introduce toate variabilele simultan ca soluţie alternativă. În acelaşi mod, puteţi să introduceţi toate variabilele simultan şi să eliminaţi treptat predictorii unul câte unul, dacă această eliminare influenţează într-o măsură foarte redusă predicţia în ansamblu.
- Puteţi, de asemenea, să introduceţi variabilele sub formă de grupuri diferite pentru analiză. Acest procedeu se numeşte regresie multiplă ierarhică şi poate fi, de pildă, ales împreună cu procedurile stepwise. Metoda folosirii grupărilor este tratată în capitolul următor.

Procesarea unei analize prin regresie multiplă stepwise este exemplificată folosind datele prezentate în tabelul 28.1, care reprezintă scorurile înregistrate de şase indivizi referitor la cele patru variabile ale reuşitei educaţionale, a capacităţii intelectuale, a motivării şcolare şi, respectiv, a interesului parental.

Având în vedere că această analiză are scop ilustrativ şi din dorinţa de a economisi spaţiu, vom introduce aceste date de 20 de ori, pentru a avea o cantitate considerabilă de informaţii de lucru. În mod evident, acest lucru *nu* se face dacă datele sunt reale. Este important să folosiţi un număr cât mai mare de participanţi sau de cazuri pentru cercetarea realizată prin regresie multiplă. Un număr de variabile de zece sau cincisprezece ori mai mare decât cel din exemplul nostru ar reprezenta o cantitate rezonabilă. Desigur, puteţi folosi un număr mai mic de participanţi atunci când obiectivul dumneavoastră este explorarea datelor.

Tabelul 28.1. *Date pentru regresia multiplă stepwise*

Educational achievement	Intellectual ability	School motivation	Parental interest
1	2	1	2
2	2	3	1
2	2	3	3
3	4	3	2
3	3	4	3
4	3	2	2

28.1. Analiza prin regresie multiplă stepwise (selecţia pas cu pas)

Pasul 1

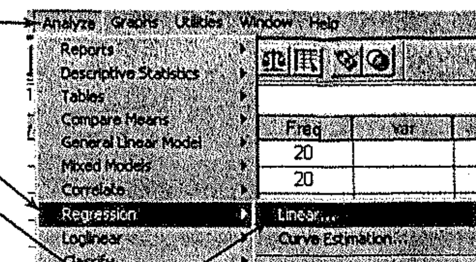
Introduceţi datele. Variabila „Freq" este folosită pentru ponderarea acestor şase cazuri de 20 de ori folosind procedura „Weight Cases...".

Salvaţi acest fişier cu date pentru a-l folosi în capitolul 29.

	Achievement	Ability	Motivation	Interest	Freq
1	1	2	1	2	20
2	2	2	3	1	20
3	2	2	3	3	20
4	3	4	3	2	20
5	3	3	4	3	20
6	4	3	2	2	20
7					

Pasul 2

Selectaţi „Analyze", „Regression" şi „Linear...".



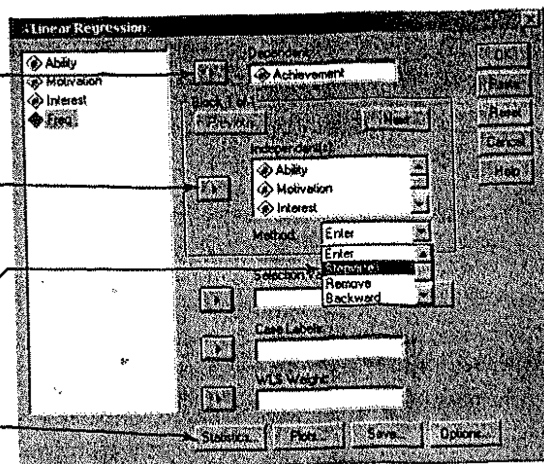
Pasul 3

Selectaţi „Achievement" şi apăsaţi butonul ▶ de lângă caseta „Dependent:" pentru a introduce această variabilă în casetă.

Selectaţi „Ability", „Motivation" şi „Interest" şi apăsaţi butonul ▶ de lângă caseta „Independent(s):" pentru a introduce aceste variabile în casetă.

Apăsaţi butonul ▼ de lângă comanda „Method:" şi alegeţi „Stepwise".

Selectaţi „Statistics...".

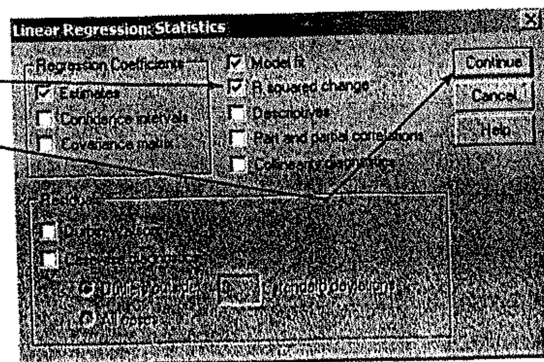


Pasul 4

Selectaţi opţiunea „R squared change".

Selectaţi „Continue".

Apăsaţi „OK" în ecranul anterior, care va reapărea.



28.2. Interpretarea output-ului

- Există o cantitate mare de informaţii în output. Regresia multiplă constituie un domeniu complex şi de aceea necesită un studiu aprofundat pentru înţelegerea tuturor implicaţiilor sale. În interpretarea rezultatelor acestei analize ne vom limita la comentarea următoarelor rezultate statistice: valoarea R multiplu, valoarea R^2, valoarea R^2 ajustat, B, beta şi R^2 corectat. (Majoritatea acestor valori sunt tratate într-o manieră accesibilă în textul statistic suplimentar – vezi *ISP*, capitolul 28.)
- În cazul regresiei multiple stepwise, fiecare nou pas este discutat sub formă de „model". În exemplul de faţă, au existat două etape semnificative. Primul pas (modelul 1) ... „Ability". Al doilea pas (modelul 2) este construit pe baza acestui predictor, adăugându-se şi un al doilea – „Motivation". În general, este bine să vă concentraţi asupra modelului cu numărul mai mare.

- Observaţi cât de rău este pus în pagină primul tabel în mod special. Dacă daţi dublu clic pe un tabel, acesta va fi încadrat într-un chenar. Pentru a deplasa orice rând în afara primului, mutaţi cursorul la rândul respectiv. Atunci când cursorul se transformă în săgeată dublă (↔), daţi clic stânga şi ţineţi butonul mouse-ului apăsat, deplasând rândul până când acesta ajunge în poziţia dorită înainte de a-l elibera. Dacă trageţi astfel de marginile coloanelor, ar trebui să obţineţi un tabel mai bun şi mai uşor de citit.

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Ability	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
2	Motivation	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

a. Variabila dependentă: Reuşita

- Al doilea tabel al output-ului, „model Summary", oferă valorile R multiplu, R^2 şi R^2 ajustat pentru cele două etape (modele). Acesta este, de fapt, un tabel al coeficienţilor de corelaţie multiplă dintre modele şi criteriu. De asemenea, acesta conţine şi statistici care indică îmbunătăţirea corelaţiei dintre modele şi date. Fiecare model din acest exemplu oferă o îmbunătăţire a corelaţiei. Aceasta se poate observa din valorile finale la nivelul cărora schimbarea corelaţiei este semnificativă atât în cazul modelului 1, cât şi în cel al modelului 2. (Ponderile de regresie – B şi beta – pot fi găsite în cel de-al patrulea tabel al output-ului, cel intitulat ...

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.701[a]	.491	.487	.689	.491	113.786	1	118	.000
2	.718[b]	.515	.507	.675	.024	5.850	1	117	.017

a. Predictors: (Constant), Ability
b. Predictors: (Constant), Ability, Motivation

Cel mai important lucru este faptul că valoarea „Sig. F Change" (valoarea corectată *F* a semnificației statisticii) indică faptul că îmbunătățirea potrivirii dintre cele două modele este semnificativă.

- Predictorul introdus în primul pas al analizei stepwise (modelul 1) este cel care are cea mai mare corelație cu respectivul criteriu. În acest exemplu, predictorul este „Ability". (Observați că litera „a" situată imediat sub tabelul „Model Summary" indică acest lucru.)
- Din moment ce în prima etapă există un singur predictor în ecuația de regresie, valoarea *R* multiplu coincide cu aceea a corelației dintre „Ability" și „Achievement" (variabila dependentă sau criteriu). În acest caz, valoarea ei este de 0,701 sau 0,70 dacă o reducem la o precizie de două zecimale.
- Valoarea R^2 reprezintă coeficientul de corelație multiplă la pătrat, care, în acest exemplu, are valoarea de 0,491 sau 0,49 dacă o reducem la două zecimale. Această cifră indică faptul că varianța criteriului este comună cu sau „explicată de" primul predictor în proporție de 49%.
- Valoarea R^2 ajustată pornește de la valoarea R^2 care a fost ajustată pentru a se potrivi dimensiunii eșantionului și numărului de predictori din ecuație. Efectul acestei ajustări este reducerea valorii R^2, astfel încât R^2 este 0,487 sau 0,49 dacă ne limităm la o precizie de două zecimale.
- A doua variabilă introdusă în ecuația de regresie este predictorul care explică, în general, a doua proporție semnificativă ca mărime a varianței criteriului. În acest exemplu, această variabilă este „Motivation".
- Valorile *R* multiplu, R^2 și R^2 ajustat pentru modelul 2 sunt 0,718, 0,515 și, respectiv, 0,507, care, rotunjite la două zecimale, devin 0,72, 0,52 și 0,51. După cum era de așteptat, aceste valori sunt mai mari decât cele din modelul 1. Acest lucru este previzibil din cauza existenței unui predictor suplimentar care contribuie la îmbunătățirea predicției.
- În Modelul 2, așadar, există două variabile („Ability" și „Motivation") care explică sau justifică în proporție de 51% varianța criteriului.
- Valoarea R^2 corectat prezentată în secțiunea „Change Statistics" din al doilea tabel prezintă creșterea proporției de varianță la nivelul variabilei criteriu („Achievement") proporțional cu predictorii care au fost introduși după primul predictor („Ability"). În acest caz, există un singur predictor („Motivation"). Acesta justifică o proporție suplimentară de 2,4% din varianța criteriului.
- Analizați tabelul intitulat „Coefficients". Uitați-vă în coloana intitulată „Beta". Prima valoare este de 0,701, pentru modelul 1. Este exact aceeași valoare cu aceea a corelației multiple din modelul 1 de mai sus. Acest lucru se întâmplă deoarece valoarea beta este coeficientul de regresie standardizat, care are aceeași valoare ca și corelația atunci când există un singur predictor. Este ca și cum toate scorurile pe care le aveți s-ar transforma în scoruri *z* înainte de începerea analizei.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	.100	.234		.428	.669
	Ability	.900	.084	.701	10.667	.000
2	(Constant)	-.167	.254		-.656	.513
	Ability	.833	.087	.649	9.561	.000
	Motivation	.167	.069	.164	2.419	.017

a. Dependent variabile: Achievement

Cea mai rapidă modalitate de accesare a output-ului regresiei multiple constă în a vă concentra asupra output-ului modelului final din tabel, cel etichetat „Coefficients". Ignorați rândul dedicat valorilor constante și concentrați-vă asupra restului rândurilor. Acestea conțin predictorii importanți ai variabilei dependente, care sunt, în mod evident, „Ability" și „Motivation", ambele elemente având o contribuție semnificativă.
Scorurile coeficienților *B* sunt greu de interpretat deoarece depind de scara de măsură folosită. Valorile beta sunt analoge coeficienților de corelație. Ambele valori beta sunt pozitive, ceea ce înseamnă că există o relație pozitivă între fiecare predictor și variabila dependentă. Abilitatea are o corelație de 0,649 cu „Achievement", iar „Motivation" contribuie cu o corelație independentă suplimentară de 0,164 cu „Achievement".
Valorile *t*, plus valorile „Sig." corespunzătoare acestora, indică o contribuție semnificativă la predicție din partea celor două variabile independente.

- Valoarea Beta este de 0,649 pentru primul predictor („Ability") și de 0,164 pentru cel de-al doilea („Motivation").
- Analiza se oprește în acest punct, din moment ce al treilea predictor („Interest") nu mai explică o proporție semnificativă a varianței criteriului. Observați că în tabelul final al output-ului intitulat „Excluded Variables", „Interest" are o valoare *t* de 0,000 și un nivel de semnificație de 1,0000. Acest lucru ne indică faptul că „Interest" este un predictor nesemnificativ al criteriului („Achievement").

28.3. Raportarea output-ului

- Există numeroase modalități de raportare a rezultatelor unei analize prin regresie multiplă stepwise. Într-un astfel de raport trebuie să includem o afirmație asemănătoare cu următoarea: „În cadrul regresiei multiple stepwise, am introdus mai întâi capacitatea intelectuală care a explicat 49% din varianța reușitei educaționale ($F_{1,117}$ = 5,85,

p = 0,017). Performanța educațională deosebită a fost asociată unei capacități intelectuale ridicate și motivației școlare".

- Uneori se prezintă și un tabel. Nu există o metodă standard de a face acest lucru, dar tabelul 28.2 este probabil la fel de clar ca majoritatea tabelelor de acest gen.

Tabelul 28.2. *Regresia multiplă stepwise a predictorilor reușitei educaționale (sunt incluși doar predictorii semnificativi)*

Variable	Multiple R	B	Standard error b	Beta	t	Significance of t
Intellectual ability	0.70	0.83	0.09	0.65	9.56	0.001
School motivation	0.72	0.17	0.07	0.16	2.42	0.05

29 Regresia multiplă ierarhică

Rezumat

- Regresia multiplă ierarhică îi permite cercetătorului să decidă în ce ordine să folosească o listă de predictori.
- Acest lucru se realizează prin asocierea predictorilor sau a grupurilor de predictori în blocuri de variabile. Computerul va efectua regresia luând fiecare bloc, în ordinea în care a fost introdus în aplicația SPSS. Așadar, această procedură constituie o metodă care obligă la analizarea variabilelor în ordinea aleasă de cercetător. În loc să lase computerul să decidă asupra criteriilor statistice, așa cum a făcut în capitolul anterior, cercetătorul hotărăște el însuși care ar trebui să fie primul predictor, al doilea și așa mai departe.
- Un bloc poate fi format dintr-un singur predictor sau dintr-un grup de predictori. Ordinea în care sunt situate variabilele în bloc nu contează, din moment ce acestea sunt tratate împreună ca grup.
- Este foarte probabil ca această ordine a blocurilor să fie aleasă din motive teoretice. Una dintre procedurile obișnuite constă în a situa în primul bloc variabilele care trebuie controlate din punct de vedere statistic. Consecința acestei decizii este faptul că variabilele de control sunt parțializate înainte ca restul blocurilor să fie analizate.
- Din momentul în care cercetătorul încearcă să obțină modele de informații, regresia multiplă poate fi variată pentru a exemplifica, de pildă, efectele introducerii blocurilor în altă ordine.
- Regresia multiplă ierarhică se folosește pentru calcularea coeficienților de corelație într-o analiză simplă a parcursului critic.

Procesarea unei analize prin intermediul regresiei multiple ierarhice este ilustrată folosind datele prezentate în tabelul 29.1, care reprezintă scorurile înregistrate în cazul a șase indivizi referitor la cele patru variabile: reușita școlară, capacitatea intelectuală, motivația școlară și interesul parental.

Am adăugat o variabilă suplimentară, clasa socială, care se măsoară pe o scară de la 1 la 5, 5 fiind cea mai înaltă clasă socială. Analiza ierarhică se folosește atunci când variabilele sunt introduse într-o ordine predeterminată de către cercetător, bazată pe motive „teoretice" și nu pe baza criteriilor statistice. Acest lucru se realizează prin ordonarea variabilelor independente în blocuri de variabile, numite „Block 1", „Block 2" etc. Un bloc poate fi alcătuit dintr-o singură variabilă independentă sau din mai multe.

Tabelul 29.1. *Date pentru regresia multiplă ierarhică*

hievement	ability	motivation	interest	class
1	2	1	2	2
2	2	3	1	1
2	2	3	3	5
3	4	3	2	4
3	3	4	3	3
4	3	2	2	2

În această analiză, vom situa în „Block 1" variabila clasei sociale („Class"), care este, practic, o variabilă demografică pe care dorim să o controlăm. „Block 2" va fi alcătuit din capacitatea intelectuală („Ability"). „Block 3" va fi format din motivația școlară („Motivation") și interesul parental („Interest"). Variabila dependentă sau criteriul pe care vrem să-l explicăm este reușita școlară („Achievement").

Practic, modelul din exemplul nostru afirmă că reușita școlară este afectată de capacitatea intelectuală, care este determinată în parte de factorii motivaționali cum ar fi motivația școlară sau interesul parental. Clasa socială este variabila pe care o controlăm în acest model din moment ce nu o considerăm a fi un factor psihologic.

Atunci când efectuăm o analiză a parcursului critic, este adeseori necesar să realizăm mai multe regresii multiple ierarhice. Acestea trebuie repetate folosindu-se blocuri diferite în ordine diferită, astfel încât diferitele modele de relații reciproce dintre acestea să poată fi cercetate.

29.1. Analiza prin regresie multiplă ierarhică

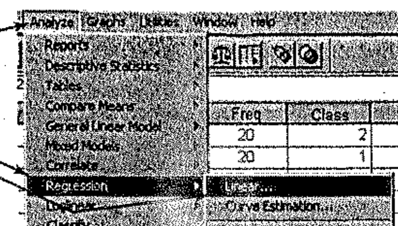
Pasul 1

Selectați fișierul de date dacă l-ați salvat. Dacă nu, introduceți datele. O a cincea variabilă, „Class", a fost introdusă în cea de-a șasea coloană.

	Achievement	Ability	Motivation	Interest	Freq	Class
1	1	2	1	2	20	2
2	2	2	3	1	20	1
3	2	2	3	3	20	5
4	3	4	3	2	20	4
5	3	3	4	3	20	3
6	4	3	2	2	20	2

Pasul 2

Selectați „Analyze", „Regression" și „Linear...".

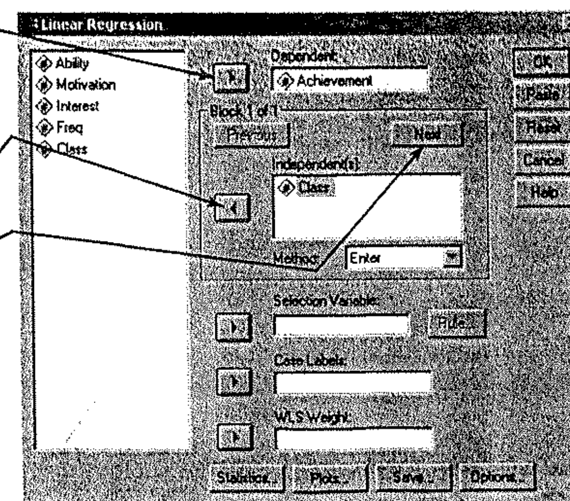


Pasul 3

Selectați „Achievement" și apăsați butonul ▶ de lângă caseta „Dependent:" pentru a introduce variabila în casetă.

Selectați „Class" și apăsați butonul ◀ de lângă caseta „Independent(s):" pentru a introduce variabila în casetă.

Selectați „Next".

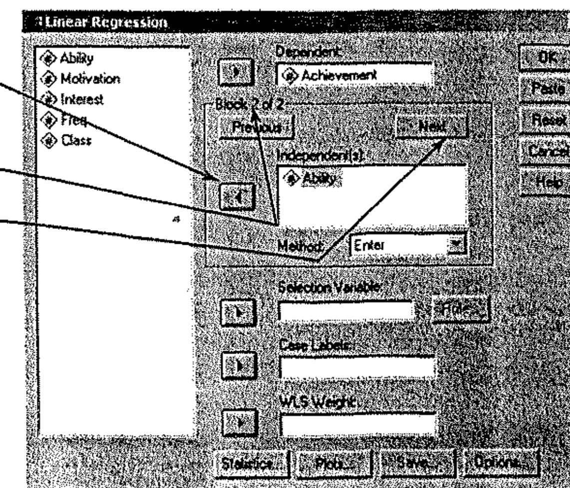


Pasul 4

Selectați „Ability" și apăsați butonul ◀ de lângă caseta „Independent(s):" pentru a introduce variabila în casetă.

Observați că acesta este „Block 2 of 2".

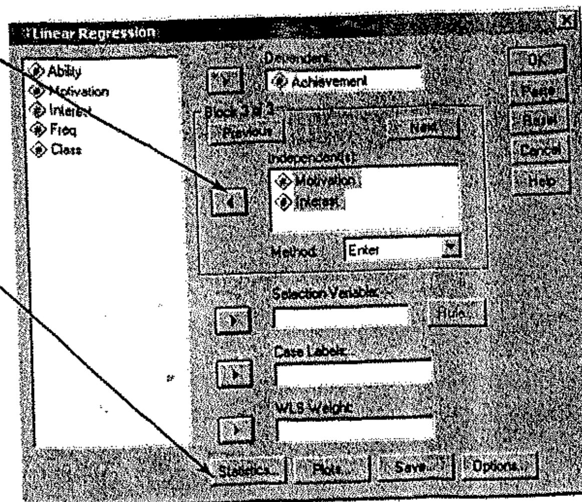
Selectați „Next".



Pasul 5

Selectați „Motivation" și „Interest" și apăsați butonul ◄ de lângă caseta „Independent(s):" pentru a introduce variabilele în casetă.

Selectați „Statistics...".

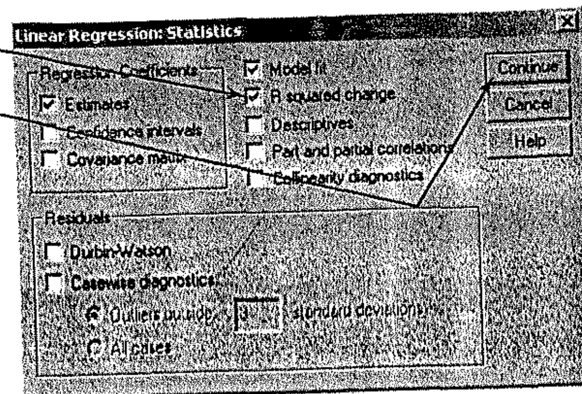


Pasul 6

Selectați „R squared change".

Selectați „Continue".

Apăsați „OK" în ecranul anterior, care reapare.



29.2. Interpretarea output-ului

- După cum se arată în al doilea tabel al output-ului intitulat „Model Summary", variabila introdusă în primul bloc este „Class". Valoarea R^2 pentru acest bloc este, practic, 0,0 (0,004), ceea ce înseamnă că variabila clasei sociale justifică în proporție de 0% varianța reușitei educaționale.
- Semnificația statistică a lui F-change cu valoarea de 0,497 pentru acest bloc sau model are semnificația de 0,482. Din moment ce această valoare este mai mare decât valoarea critică de 0,05, aceasta înseamnă că ecuația de regresie din această primă etapă nu explică într-o proporție semnificativă varianța variabilei reușita educațională/școlară.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.066ᵃ	.004	-.004	.963	.004	.497	1	118	.482
2	.714ᵇ	.509	.501	.679	.505	120.333	1	117	.000
3	.769ᶜ	.591	.577	.626	.082	11.500	2	115	.000

a. Predictors: (Constant), Clasa.
b. Predictors: (Constant), Clasa, Ability.
c. Predictors: (Constant), Clasa, Ability, Motivation, Interest.

- Variabila introdusă în cel de-al doilea bloc este „Ability". Valoarea R^2 ajustat pentru acest bloc sau model este de 0,501, ceea ce înseamnă că variabila capacitate intelectuală, împreună cu cea a clasei sociale, justifică în proporție de 50,1% varianța reușitei educaționale (școlară).
- Semnificația statistică a Raportului F pentru acest bloc este de 0,000, ceea ce înseamnă că este mai mică decât 0,001. Din moment ce această valoare este mult mai mică decât valoarea critică de 0,05, vom trage concluzia că primii doi pași ai ecuației de regresie explică într-o proporție semnificativă varianța reușitei educaționale.
- Variabilele introduse în cel de-al treilea și ultimul bloc sunt „Motivation" și „Interest". Valoarea R^2 ajustat pentru acest bloc este de 0,557, ceea ce înseamnă că toate cele patru variabile justifică în proporție de 55,7% varianța reușitei educaționale.
- Raportul F pentru acest bloc este de 0,000. Din moment ce această valoare este mult mai mică decât valoarea critică de 0,05, înseamnă că primii trei pași ai ecuației de regresie justifică într-o proporție semnificativă varianța reușitei educaționale.
- Cea mai simplă interpretare a output-ului se poate realiza prin examinarea celui de-al patrulea tabel al output-ului, intitulat „Coefficients". Coloanele „Beta" și „Sig." (ale valorii t) sunt deosebit de utile. Acestea ne comunică faptul că valoarea corelației (beta) dintre variabilele „Class" și „Achievement" este -0,439 atunci când sunt luați în considerare ceilalți predictori. Această corelație este semnificativă la nivel de 0,000, ceea ce înseamnă mai puțin de 0,001. Acest coeficient este semnificativ acum pentru că variabilele „Ability" și „Interest" suprimă coeficientul de ordin zero dintre „Class" și „Achievement". În condițiile în care controlăm variabila clasei sociale în blocul 1, corelația dintre „Ability" și „Achievement" este de 0,730. Această valoare este semnificativă la nivel de 0,000. În cele din urmă, în condițiile în care controlăm variabilele „Class" și „Ability", corelațiile pentru fiecare dintre variabilele din blocul 3 (motivația școlară și interesul parental) și reușita educațională („Achievement") sunt date separat.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	2.369	.205		11.543	.000
	Class	.046	.065	.065	.705	.482
2	(Constant)	.250	.241		1.036	.302
	Class	–.100	.048	–.140	–2.082	.040
	Ability	.950	.087	.740	10.970	.000
3	(Constant)	–.562	.284		–1.984	.050
	Class	–.313	.068	–.439	–4.615	.000
	Ability	.938	.084	.730	11.180	.000
	Motivation	.187	.068	.185	2.769	.007
	Interest	.438	.130	.314	3.374	.001

a. Dependent Variable : Achievement

Cea mai rapidă modalitate de accesare a output-ului este de a vă concentra asupra celui de-al treilea model din datele de mai sus. Output-ul vă indică faptul că atunci când toate variabilele „Class", „Ability", „Motivation" şi „Interest" sunt introduse, există o asociere semnificativă între fiecare dintre ele şi variabila dependentă „Achievement". Clasa socială are o relaţie negativă, în timp ce restul variabilelor au o relaţie pozitivă.

29.3. Raportarea output-ului

- Există diferite moduri de raportare a rezultatelor analizei prin regresie multiplă ierarhică. Într-un astfel de raport, de obicei se descriu procentele de varianţă, explicate de fiecare set sau bloc de predictori (după valoarea R^2).
- O modalitate de raportare a acestor rezultate ar fi următoarea: „Într-o regresie multiplă ierarhică, predictorii potenţiali ai reuşitei au fost asociaţi în blocuri. Clasa socială a fost introdusă prima, apoi capacitatea intelectuală a fost adăugată în al doilea bloc, iar motivaţia şcolară şi interesul parental au fost adăugate în ultimul bloc. Modelul final a indicat faptul că variabila clasei sociale este un predictor negativ (B = – 0,31), capacitatea intelectuală unul pozitiv (B = 0,94), iar motivaţia şcolară şi interesul parental sunt şi ei predictori pozitivi (B = 0,19 şi 0,44). Toţi predictorii au fost semnificativi la nivel de 1%".
- De asemenea, trebuie rezumată şi ecuaţia de regresie, ca în tabelul 29.2.

Tabelul 29.2. *Regresia multiplă ierarhică a predictorilor reuşitei educaţionale*

Blocks	*B*	Standard error *B*	Beta
Block 1			
Social class	–0.31	0.07	–0.44*
Block 2			
Intellectual ability	0.94	0.08	0.73*
Block 3			
School motivation	0.19	0.07	0.19*
Parental interest	0.44	0.13	0.31*

* Semnificativ la nivel de 0,01.

30 Analiza validităţii itemilor şi acordul interevaluare

Rezumat

- Validitatea este o chestiune complexă, deoarece acest termen se referă la o gamă de concepte şi măsuri foarte diferite. Este uşor să le încurci între ele.
- Coeficientul alpha de validitate a itemilor şi metoda split-half evaluează consistenţa internă a itemilor unui chestionar – adică, felul în care itemii tind să măsoare acelaşi lucru.
- Metoda split-half pentru studiul validităţii testului din aplicaţia SPSS se referă la corelaţia dintre scorurile obţinute pe baza primei jumătăţi de itemi incluşi în chestionar şi cele obţinute pe baza celei de-a doua jumătăţi a itemilor chestionarului. Această corelaţie poate fi ajustată statistic pentru a se menţine lungimea chestionarului iniţial.
- Coeficientul alpha nu este altceva decât o medie a tuturor fidelităţilor chestionarului obţinute prin metoda split-half pentru un chestionar şi de aceea ar putea fi de preferat, din moment ce nu depinde de ordinea itemilor în chestionar. Coeficientul alpha poate fi folosit ca modalitate de scurtare a chestionarului, menţinând sau îmbunătăţind în acelaşi timp validitatea sa internă.
- Acordul interevaluare (măsurată aici de coeficientul kappa) este, practic, o măsură a concordanţei dintre aprecierile a doi evaluatori diferiţi. Aşadar, această metodă este deosebit de folositoare pentru evaluarea codificărilor sau aprecierilor efectuate de către „experţi" ale informaţiilor obţinute pe baza întrebărilor deschise; cu alte cuvinte, se referă la cuantificarea datelor calitative. Acest procedeu implică măsura în care cei doi evaluatori au fost întru totul de acord asupra evaluărilor lor, în comparaţie cu acordul prevăzut de probabilitate. Observaţi că acest aspect este diferit de corelaţia dintre evaluatori care nu necesită un acord *perfect* pentru atingerea unui nivel ridicat al corelaţiei, ci doar o creştere relativă a evaluărilor în cazul ambilor evaluatori.
- Alte forme de validitate, cum ar fi consistenţa unei măsuri luate în două momente diferite (fidelitatea test-retest), pot fi evaluate prin simpla folosire a coeficientului de corelaţie (capitolul 7).

30.1. Coeficientul alpha de validitate a itemilor

Răspunsurile date de zece persoane referitor la cei patru itemi ai unui chestionar sunt ilustrate în tabelul 30.1. Aceste informaţii vor fi folosite pentru ilustrarea a două metode de stabilire a fidelităţii itemilor, cunoscute sub numele de coeficientul alpha de fidelitate şi metoda split-half.

Tabelul 30.1. *Informaţii referitoare la 10 cazuri, pe baza unui chestionar cu patru itemi*

Cases	Item 1	Item 2	Item 3	Item 4
1	1	3	5	6
2	2	1	1	2
3	1	1	1	1
4	5	2	4	2
5	6	4	3	2
6	5	4	5	6
7	4	5	3	2
8	2	1	2	1
9	1	2	1	1
10	1	1	2	2

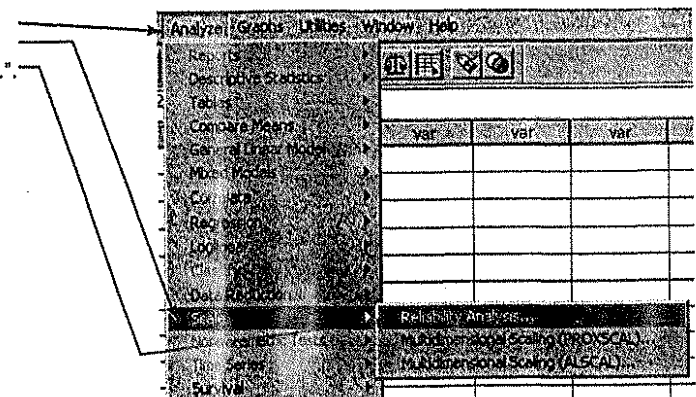
Pasul 1

Introduceţi datele.

	item1	item2	item3	item4
1	1	3	5	6
2	2	1	1	2
3	1	1	1	1
4	5	2	4	2
5	6	4	3	2
6	5	4	5	6
7	4	5	3	2
8	2	1	2	1
9	1	2	1	1
10	1	1	2	2

Pasul 2

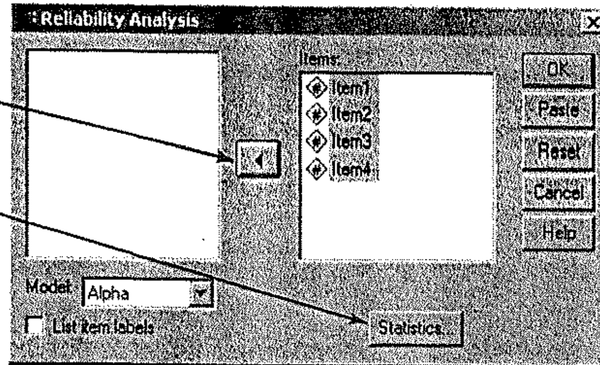
Selectaţi „Analyze", „Scale" şi „Reliability Analysis...".



Pasul 3

Selectați cei patru itemi separat sau împreună, iar apoi apăsați butonul ◄ pentru a le introduce în caseta „Items:".

Selectați „Statistics...".

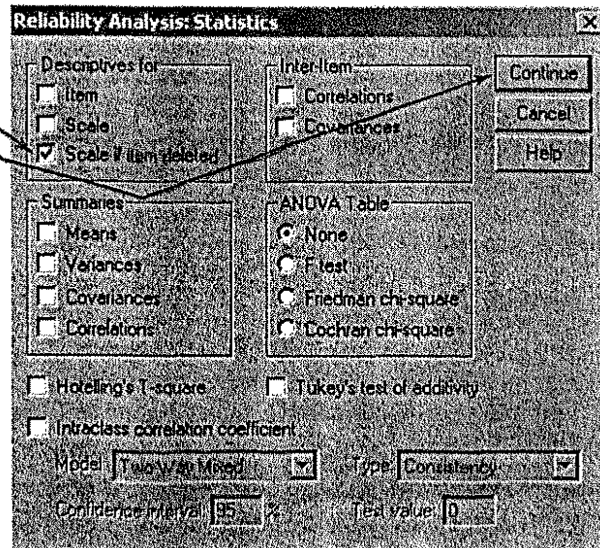


Pasul 4

Selectați opțiunea „Scale if item deleted".

Selectați „Continue".

Apăsați „OK" în ecranul anterior care va reapărea.



30.2. Interpretarea output-ului

Primul tabel ilustrează numărul de cazuri pe care se bazează analiza, adică 10.

Case Processing Summary

		N	%
Cases	Valid	10	100.0
	Excluded[a]	0	.0
	Total	10	100.0

a. Ştergerea de pe listă, pe baza tuturor variabilelor din procedură.

Al doilea tabel reprezintă coeficientul alpha de validitate a celor patru itemi şi are valoarea de 0,811, valoare care se rotunjeşte la un număr cu două zecimale egal cu 0,81.

Reliability Statistics

Cronbach's Alpha	N of Items
.811	4

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Item 1	7.60	18.933	.490	.840
Item 2	8.00	19.556	.718	.731
Item 3	7.70	17.789	.842	.671
Item 4	7.90	18.767	.547	.806

Acesta este coeficientul alpha de validitate a unei scale din care a fost eliminat unul dintre itemi. Din această coloană putem vedea că, dacă îndepărtăm primul item („Item 1"), coeficientul alpha de validitate al celor trei itemi ai scalei care rămân creşte uşor de la 0,811 la 0,840. Din moment ce aceasta este o schimbare foarte mică, probabil că este mai bine să păstrăm şi itemul 1.

30.3. Raportarea output-ului

O modalitate de raportare a rezultatelor acestei analize ar fi următoarea: „Coeficientul alpha de validitate a scalei cu patru itemi are valoarea de 0,81, ceea ce demonstrează că scala are un nivel bun de fidelitate". Un coeficient alpha cu o valoare mai mare sau egală cu 0,80 este considerat satisfăcător.

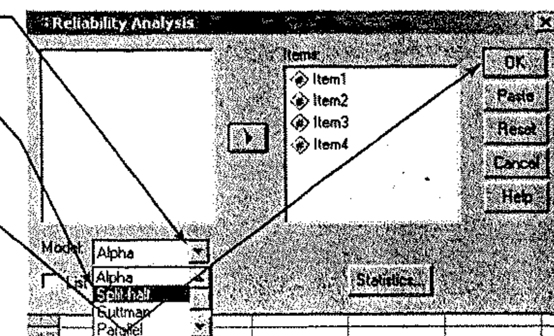
30.4. Metoda split-half

Datele prezentate mai sus sunt folosite din nou şi pentru această analiză.

Apăsaţi pe butonul ▼ din fereastra „Model:" şi alegeţi opţiunea „Split-half".

Apăsaţi „OK".

(Dacă doriţi, puteţi deselecta opţiunea „Scale if item deleted" din meniul „Statistics...".)



30.5. Interpretarea output-ului

Ultimul rând din al doilea tabel ilustrează coeficientul de validitate obținut prin testul Guttman în cazul celor patru itemi ca având valoarea de 0,646, care se rotunjește la un număr cu două zecimale egal cu 0,65.

Reliability Statistics

Cronbach's Alpha	Part 1	Value	.777
		N of Items	2[a]
	Part 2	Value	.904
		N of Items	2[b]
	Total N of Items		4
Correlation Between Forms			.477
Spearman-Brown Coefficient	Equal Length		.646
	Unequal Length		.646
Guttman Split-Half Coefficient			.646

a. Itemii sunt: itemul 1, itemul 2.
b. Itemii sunt: itemul 3, itemul 4.

30.6. Raportarea output-ului

O modalitate de raportare a rezultatelor acestei analize ar fi următoarea: „Coeficientul de fidelitate obținut prin metoda split-half în cazul scalei cu patru itemi este de 0,65, ceea ce indică faptul că fidelitatea scalei este una moderată".

30.7. Corelația intraclasă (coeficientul de concordanță kappa)

Coeficientul kappa este folosit pentru măsurarea concordanței dintre doi evaluatori, având în vedere acordul estimat prin probabilitate. Vom ilustra calcularea acestui coeficient folosind datele din tabelul 30.2, care conține aprecierile făcute de un psiholog criminalist și de un psihiatru asupra a 12 persoane vinovate de infracțiuni sexuale, evaluând măsura în care acestea nu reprezintă pentru public nici un risc (1), constituie un risc moderat (2) sau un risc mare (3).

Tabelul 30.2. *Evaluările riscului prezentat de 12 infractori, efectuate de către doi specialiști*

Sex offenders	Forensic psychologist	Psychiatrist
1	3	3
2	3	3
3	3	3
4	1	1
5	1	2
6	3	3
7	2	3
8	3	3
9	2	3
10	3	3
11	3	3
12	3	3

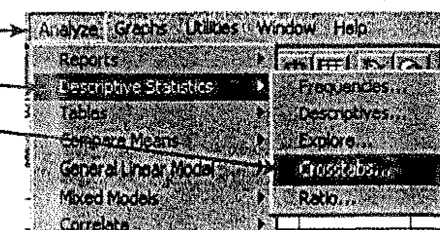
Pasul 1

Introduceți datele.

	Psychologist	Psychiatrist
1	3	3
2	3	3
3	3	3
4	1	1
5	1	2
6	3	3
7	2	3
8	3	3
9	2	3
10	3	3
11	3	3
12	3	3

Pasul 2

Selectați „Analyze", „Descriptive Statistics" și „Crosstabs...".

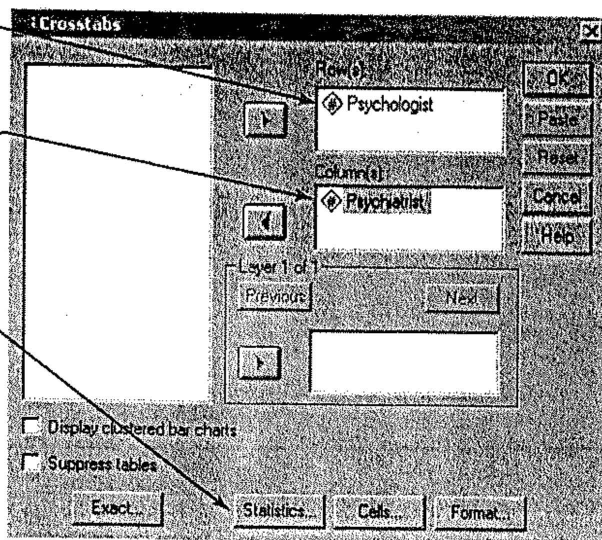


Pasul 3

Selectați „Psychologist" și apăsați butonul ▸ de lângă caseta „Row(s):" pentru a introduce informația în casetă.

Selectați „Psychiatrist" și apăsați butonul ◂ de lângă caseta „Column(s):" pentru a introduce informația în casetă.

Selectați „Statistics...".

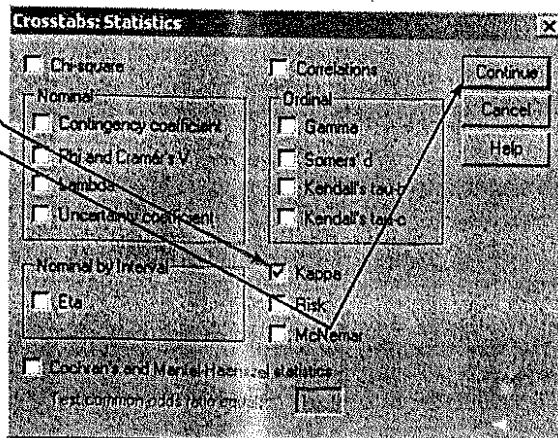


Pasul 4

Selectați „Kappa".

Selectați „Continue".

Apăsați „OK" în ecranul anterior, care va reapărea.



30.8. Interpretarea output-ului

În al doilea tabel al output-ului, informațiile au fost aranjate într-un format 2×2. Numărul de cazuri asupra cărora psihologul criminalist și psihiatrul sunt de aceeași părere sunt ilustrate în celule de pe diagonala acestui tabel. Aceste valori sunt 1 pentru evaluarea 1, 0 pentru evaluarea 2 și 8 pentru evaluarea 3.

Psychologist * Psychiatrist Crosstabulation

Count

		Psychiatrist			Total
		1	2	3	
Psychologist	1	1	1	0	2
	2	0	0	2	2
	3	0	0	8	8
Total		1	1	10	12

În al treilea tabel, kappa apare ca având valoarea de 0,400. Deși coeficientul kappa are o semnificație statistică $p = 0{,}046$, această valoare indică doar un nivel moderat de acord.

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Measure of Agreement	Kappa	.400	.219	2.000	.046
N of Valid Cases		12			

a. Neluând în considerare ipoteza nulă.
b. Folosind eroarea standard asimptotică, luând în cosiderare ipoteza nulă.

Observați că valoarea kappa le permite evaluatorilor să folosească aceeași evaluare de mai multe ori. Aceasta *nu* reprezintă o măsură a proporției corelației dintre cei doi.

30.9. Raportarea output-ului

O modalitate de raportare a rezultatelor acestei analize ar fi următoarea: „Coeficientul kappa al corelației dintre evaluările pshilogului criminalist și ale psihiatrului a fost de 0,40, ceea ce indică o valoare moderată a corelației".

31 Analiza logliniară

Rezumat

- Analiza logliniară se folosşte pentru analizarea tabelelor de contingenţă care conţin trei sau mai multe variabile. Aşadar, poate fi considerată a fi o continuare a Testului chi-square discutat în capitolul 14.
- Scopul acestei analize este acela de a determina care dintre variabile şi interacţiunile acestora explică (sau redau) cel mai bine frecvenţele observate din tabel. Variabilele şi interacţiunile acestora considerate pe cont propriu sau în combinaţie sunt cunoscute sub numele de modele.
- Statisticile efectuate prin teste de tipul *goodness-of-fit* sunt folosite pentru evaluarea gradului de corespondenţă dintre model şi date. Semnificaţia statistică indică faptul că modelul care este examinat nu reuşeşte să justifice în totalitate frecvenţele observate. Lipsa semnificaţiei statistice înseamnă că modelul care este analizat se potriveşte cu frecvenţele observate. Dacă există mai mult de un model care se potriveşte cu datele, modelul care are cele mai puţine variabile şi interacţiuni este considerat a fi cel mai simplu şi, de aceea, poate fi modelul preferat. Raportul de probabilitate chi-square este folosit ca metodă statistică de testare.

Procesarea unei analize logliniare este exemplificată folosind datele din tabelul 31.1. Acest tabel conţine informaţii referitoare la frecvenţa abuzului sexual şi fizic, în cazul unor pacienţi cu afecţiuni psihice, dintre care 140 sunt femei, iar 160 – bărbaţi. Pentru analizarea unui tabel cu date ca acesta cu ajutorul aplicaţiei SPSS, mai întâi trebuie să introducem datele în „Data Editor" şi să ponderăm celulele după frecvenţele cazurilor pe care le conţin.

Tabelul 31.1. *O contingenţă 3-way care ilustrează relaţia dintre sex, abuz sexual şi abuz fizic, în cazul unui eşantion de pacienţi ai unui spital psihiatric*

Sexual abuse	Physical abuse	Sex		Margin totals
		Female	Male	
Sexually abused	Physical abuse	20	30	50
	No physical	40	25	65
Not sexually abused	Physical abuse	35	55	90
	No physical	45	50	95
Margin totals		140	160	300

31.1. Analiza logliniară

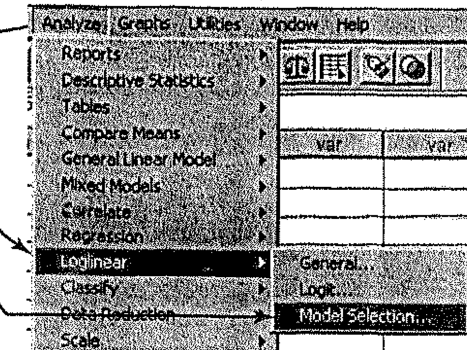
Pasul 1

Introduceţi datele. Ponderaţi cazurile după valoarea „Freq.".

	Sexual	Physical	Sex	Freq
1	1	1	1	20
2	1	1	2	30
3	1	2	1	40
4	1	2	2	25
5	2	1	1	35
6	2	1	2	55
7	2	2	1	45
8	2	2	2	50

Pasul 2

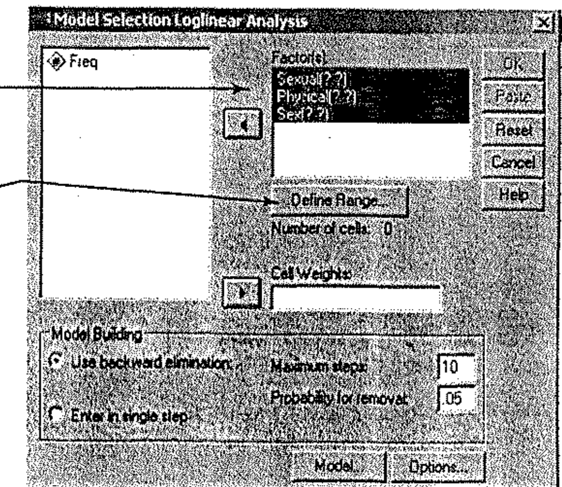
Selectaţi „Analyze", „Loglinear" şi „Model Selection...".



Pasul 3

Selectaţi separat sau împreună factorii „Sexual", „Physical" şi „Sex" şi apăsaţi butonul ◄ de lângă caseta „Factor(s):" pentru a introduce elementele în casetă.

Selectaţi „Define Range". Din moment ce toate cele trei variabile au acelaşi interval, acesta poate fi definit odată pentru toate.

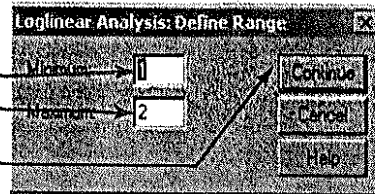


Pasul 4

Scrieţi „1" în caseta „Minimum:".
Scrieţi „2" în caseta „Maximum:".

Selectaţi „Continue".

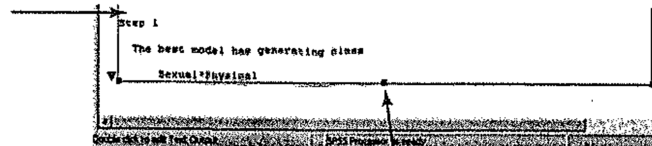
Apăsaţi „OK" în ecranul anterior, care va reapărea.



31.2. Interpretarea output-ului

Pentru a desfăşura output-ul complet prezentat mai jos, daţi clic în fereastra „Viewer" care creează un cadru în jurul output-ului. Săgeata îndreptată în jos, spre colţul din stânga jos al output-ului, indică faptul că mai aveţi de văzut o parte din output. Aşezaţi cursorul pe micul pătrat din centrul liniei din josul paginii şi trageţi această linie în jos până când se afişează tot output-ul.

Daţi clic în output îndepărtându-vă de margini, pentru a forma conturul.



Deplasaţi cursorul aici, daţi clic şi trageţi linia în jos până când ajungeţi la finalul output-ului. Aveţi nevoie de oarecare exerciţiu pentru a face acest lucru.

Output-ul extins este prezentat dacă faceţi acest lucru. De cele mai multe ori, este mai bine să-l ignoraţi. Modelul final este cel mai important.

```
*********** HIERARCHICAL LOG LINEAR ***********
DATA    Information
          8 unweighted cases accepted.
          0 cases rejected because of out-of-range factor values.
          0 cases rejected because of missing data.
        300 weighted cases will be used in the analysis.
FACTOR    Information
   Factor    Level Label
   SEXUAL        2 Sexual abuse
   PHYSICAL      2 Physical abuse
   SEX           2 Sex
*********** HIERARCHICAL LOG LINEAR ***********
```



```
--------------------------------------------------------------------------------
DESIGN 1 has generating class
   SEXUAL*PHYSICAL*SEX
  Note: For saturated models .500 has been added to all observed
  cells.
  This value may be changed by using the CRITERIA = DELTA subcommand.
The Iterative Proportional Fit algorithm converged at iteration 1.
The maximum difference between observed and fitted marginal totals is
.000 and the convergence criterion is .250
--------------------------------------------------------------------------------
Observed, Expected Frequencies and Residuals.
   Factor     Code          OBS count     EXP count     Residual      Std Resid
SEXUAL        Sexually
 PHYSICAL      Physical
  SEX           Females          20.5          20.5          .00            .00
  SEX           Males            30.5          30.5          .00            .00
 PHYSICAL      Not phys
  SEX           Females          40.5          40.5          .00            .00
  SEX           Males            25.5          25.5          .00            .00
SEXUAL        Not sexu
 PHYSICAL      Physical
  SEX           Females          35.5          35.5          .00            .00
  SEX           Males            55.5          55.5          .00            .00
 PHYSICAL      Not phys
  SEX           Females          45.5          45.5          .00            .00
  SEX           Males            50.5          50.5          .00            .00
--------------------------------------------------------------------------------
Goodness-of-fit test statistics
  Likelihood ratio chi square =    .00000    DF = 0    P = 1.000
          Pearson chi square =    .00000    DF = 0    P = 1.000
--------------------------------------------------------------------------------
*********** HIERARCHICAL LOG LINEAR ***********
Tests that K-way and higher order effects are zero.
        K   DF    L.R. Chisq    Prob   Pearson Chisq    Prob   Iteration
        3    1         1.185   .2764           1.181   .2772           3
        2    4         9.680   .0462          10.013   .0402           2
        1    7        28.834   .0002          28.000   .0002           0
--------------------------------------------------------------------------------
Tests that K-way effects are zero.
        K   DF    L.R. Chisq    Prob   Pearson Chisq    Prob   Iteration
        1    3        19.154   .0003          17.987   .0004           0
        2    3         8.495   .0368           8.833   .0316           0
        3    1         1.185   .2764           1.181   .2772           0
*********** HIERARCHICAL LOG LINEAR **********
Backward Elimination (p = .050) for DESIGN 1 with generating class
   SEXUAL*PHYSICAL*SEX
  Likelihood ratio chi square =    .00000    DF = 0    P = 1.000
```

```
If Deleted Simple Effect is  DF    L.R. Chisq Change    Prob    Iter
 SEXUAL*PHYSICAL*SEX          1               1.185    .2674      3
Step 1
 The best model has generating class
   SEXUAL*PHYSICAL
   SEXUAL*SEX
   PHYSICAL*SEX
Likelihood ratio chi square =    1.18471     DF = 1     P = .276
---------------------------------------------------------------------
If Deleted Simple Effect is  DF    L.R. Chisq Change    Prob    Iter
 SEXUAL*PHYSICAL*SEX          1                .454    .5005      2
 SEXUAL*SEX                   1               1.963    .1612      2
 PHYSICAL*SEX                 1               5.461    .0194      2
Step 2
 The best model has generating class
   SEXUAL*SEX
   PHYSICAL*SEX
Likelihood ratio chi square =    1.63849     DF = 2     P = .441
---------------------------------------------------------------------
If Deleted Simple Effect is  DF    L.R. Chisq Change    Prob    Iter
 SEXUAL*SEX                   1               2.272    .1317      2
 PHYSICAL*SEX                 1               5.770    .0163      2
************ HIERARCHICAL LOG LINEAR ***********
Step 3
 The best model has generating class
   PHYSICAL*SEX
   SEXUAL
Likelihood ratio chi square =    3.91036     DF = 3     P = .271
---------------------------------------------------------------------
If Deleted Simple Effect is  DF    L.R. Chisq Change    Prob    Iter
 PHYSICAL*SEX                 1               5.770    .0163      2
 SEXUAL                       1              16.485    .0000      2
 Step 4
 The best model has generating class
   PHYSICAL*SEX
   SEXUAL
Likelihood ratio chi square =    3.91036     DF = 3     P = .271
```



```
---------------------------------------------------------------------
************HIERARCHICAL LOG LINEAR ***********
The final model has generating class
  PHYSICAL*SEX
  SEXUAL
The Iterative Proportional Fit algorithm converged at iteration 0.
The maximum difference between observed and fitted marginal totals is
.000 and the convergence criterion is .250
---------------------------------------------------------------------
Observed, Expected Frequencies and Residuals.
  Factor        Code         OBS count   EXP count   Residual   Std Resid
SEXUAL        Sexually
 PHYSICAL      Physical
  SEX           Females        20.0        21.1       -1.08        -.24
  SEX           Males          30.0        32.6       -2.58        -.45
 PHYSICAL      Not phys
  SEX           Females        40.0        32.6        7.42        1.30
  SEX           Males          25.0        28.8       -3.75        -.70
SEXUAL        Not sexu
 PHYSICAL      Physical
  SEX           Females        35.0        35.9        1.08         .19
  SEX           Males          55.0        52.4        2.58         .36
 PHYSICAL      Not phys
  SEX           Females        45.0        52.4       -7.42       -1.02
  SEX           Males          50.0        46.3        3.75         .55
---------------------------------------------------------------------
Goodness-of-fit test statistics
  Likelihood ratio chi square =   3.91036    DF = 3    P = .271
          Pearson chi square =   3.95320    DF = 3    P = .267
---------------------------------------------------------------------
```

Acesta este modelul final. Mai jos sunt afișate frecvențele pentru datele observate, iar dedesubt, cele pentru datele anticipate. Valorile reziduale reprezintă diferențele dintre datele observate și cele preconizate de model.

- Au fost folosite două statistici pentru testarea gradului de corespondență a diferitelor modele. Acestea sunt Raportul de probabilitate chi-square și Raportul chi-square Pearson. Raportul de probabilitate chi-square este testul folosit cel mai des, deoarece are avantajul de a fi liniar, ceea ce permite adăugarea și eliminarea unor valori chi-square.
- Raportul de probabilitate chi-square pentru modelul saturat sau complet este primul prezentat și valoarea sa este de 0,00000, care are o probabilitate de 1,000. Cu alte cuvinte, modelul saturat oferă o potrivire perfectă a frecvențelor observate și, în consecință, este nesemnificativ. Modelul saturat este format, în acest caz, din trei efecte principale, trei interacțiuni 2-way și o interacțiune 3-way. În general, modelul saturat conține toate efectele principale și interacțiunile.
- Totuși, modelul saturat conține *toate* componentele, indiferent dacă acestea contribuie sau nu la explicarea variației de la nivelul datelor observate. Așadar, este necesară eliminarea componentelor pe rând, pentru a vedea dacă astfel se micșorează gradul de corespondență a modelului. Dacă se întâmplă acest lucru, componenta modelului se menține în modelul final.
- SPSS începe cu modelul complet și elimină pe rând fiecare efect, pentru a determina care dintre efecte produce schimbarea cea mai puțin semnificativă la nivelul raportului de probabilitate chi-square.

- Modelul cu gradul cel mai înalt de corespondență este prezentat ultimul. În exemplul nostru, acesta conține interacțiunea dintre abuzul fizic și sex și efectul principal al abuzului sexual. Acest model are un raport de probabilitate chi-square cu o valoare de 3,91 (rotunjit la un număr cu două zecimale), 2 grade de libertate și un nivel de probabilitate de 0,271. Cu alte cuvinte nu este semnificativ, ceea ce înseamnă că datele observate pot fi reproduse cu aceste două efecte.
- Pentru a interpreta aceste două efecte, trebuie să prezentăm datele în condițiile unui tabel one-way pentru abuzul sexual și ale unuia 2-way pentru abuzul fizic și pentru sex. Putem face acest lucru folosind raportul chi-square... pentru tabelul one-way și crosstabs... pentru tabelul bilateral. Aceste două tabele sunt prezentate mai jos. Tabelul one-way demonstrează că există mai mulți pacienți cu afecțiuni psihice care nu au fost abuzați sexual decât cei care au fost abuzați sexual. Tabelul 2-way indică faptul că în cazul bărbaților prezența abuzurilor fizice este mai probabilă decât în cazul femeilor.

Physical abuse * Sex Crosstabulation

			Sex		Total
			Females	Males	
Physical abuse	Physically abused	Count	55	85	140
		Expected Count	65.3	74.7	140.0
		Residual	-10.3	10.3	
	Not physically abused	Count	85	75	160
		Expected Count	74.7	85.3	160.0
		Residual	10.3	-10.3	
Total		Count	140	160	300
		Expected Count	140.0	160.0	300.0

Sexual abuse

	Observed N	Expected N	Residual
Sexually abused	115	150.0	-35.0
Not sexually abused	185	150.0	35.0
Total	300		

- Putem vedea care este contribuția fiecărei componente la modelul final. Chiar înainte de etapa finală (pasul 4 în acest exemplu), există un tabel mic intitulat „If Deleted Simple Effect is". Acest tabel conține o coloană numită „L.R. Chisq Change". Aceste informații indică, practic, ajustarea (reducerea) gradului de corespondență al Raportului chi-square atunci când este îndepărtată fiecare componentă. Așadar „Physical*Sex" are un raport de probabilitate chi-square corectat de 5.770, valoare care este semnificativă (0,0163). „Sexual" are o valoare de 16,485 care are o semnificație mare (0,0000). În mod evident, aceste două efecte nu pot fi eliminate din model, din cauza contribuției lor semnificative.
- Într-un model ierarhic, componentele unei interacțiuni pot fi semnificative. Din moment ce „Physical*Sex" are o contribuție semnificativă la model, „Physical" și „Sex" pot fi și ele efecte principale semnificative. Selectați „Model" în caseta de dialog „Model" de la nivelul selecției analizei logliniare (pasul 3 din secțiunea 31.1). Fereastra care apare vă va permite să testați efectele principale prin specificarea modelelor care conțin doar aceste efecte principale specifice.

31.3. Raportarea rezultatelor

O modalitate de raportare a rezultatelor obținute ar fi următoarea: „S-a efectuat o analiză de frecvență 3-way, pentru elaborarea unui model liniar ierarhic al abuzului fizic și sexual la pacienți cu afecțiuni psihice atât bărbați, cât și femei. Prin metoda eliminării pas cu pas, s-a obținut un model care conține efectul principal al abuzului sexual și efectul de interacțiune dintre abuzul fizic și sex. Modelul are un raport de probabilitate $\chi^2(3) = 3{,}91$, $p = 0{,}27$, valori care indică un grad mare de corespondență între frecvențele observate și frecvențele anticipate generate de model. Aproximativ 38% din pacienții cu afecțiuni psihice au fost abuzați sexual. Aproximativ 53% din bărbați au fost abuzați fizic, în comparație cu un procent de 39% în cazul femeilor".

32 Regresia logistică multinomială

Rezumat

- Regresia logistică este o formă de regresie multiplă (vezi capitolul 28). Aceasta identifică variabilele care disting la nivel colectiv cazurile ce aparţin diferitelor categorii ale unei variabile nominale (sau categoriale). De exemplu, poate fi folosită pentru identificarea caracteristicilor distinctive ale studenţilor la psihologie, sociologie sau fizică.
- În altă ordine de idei, regresia logistică identifică grupurile de variabile care clasifică cu precizie persoanele în funcţie de apartenenţa acestora la diferitele categorii ale unei variabile nominale.
- Regresia logistică binomială se foloseşte dacă există doar două categorii de variabile care trebuie prevăzute. Această situaţie este tratată în capitolul 33. Regresia logistică multinomială se foloseşte atunci când există trei sau mai multe categorii.
- Predictorii (adică variabilele independente) pot fi variabile numerice, variabile nominale (sau categoriale) sau pot fi alcătuiţi dintr-o combinaţie a celor două categorii.
- Cei mai buni predictori din categoria căreia îi aparţine un caz au ponderi-*b* (sau ponderi de regresie) semnificative. Această practică este foarte asemănătoare cu procedurile mai cunoscute de analiză prin regresie multiplă descrise în capitolul 28.
- Valoarea lui *b* din regresia logistică se aplică, de fapt, la logaritmul natural al unei valori numite *raportul şanselor*, care reprezintă raportul frecvenţelor a două rezultate alternative. Acest logaritm este cunoscut şi sub numele de *logit*. De aici, termenul de regresie logistică. Raportul şanselor este pur şi simplu probabilitatea de apartenenţă la o categorie şi nu la alta. Majoritatea cercetătorilor au rareori nevoie să calculeze aceste valori, astfel încât valoarea logit are mai curând importanţă conceptuală decât practică.
- Este mult mai important să înţelegem conceptul de *variabilă de substituţie*. Aceasta este o metodă care ne dă posibilitatea de a trata numeric o variabilă nominală. Dacă variabila nominală are doar *două* categorii, atunci acestea pot fi codificate numeric prin valorile 0 şi 1.
- Totuşi, dacă respectiva categorie nominală are *trei* sau mai multe categorii, procesul devine ceva mai complex. Practic, datele sunt codificate pentru a exprima prezenţa sau absenţa fiecăreia dintre cele *trei* (sau mai multe) categorii. În realitate, se creează trei (sau mai multe) variabile noi. Aşadar, dacă cele trei categorii se numesc A, B şi C, se mai creează alte trei variabile noi:
  (i) Individul aparţine Categoriei A (sau nu).
  (ii) Individul aparţine Categoriei B (sau nu).
  (iii) Individul aparţine Categoriei C (sau nu).
  Fiecare dintre aceste variabile reprezintă câte o variabilă de substituţie.

- Totuşi, întotdeauna se exclude din analiză o variabilă de substituţie din acest set. Nu contează care anume. Motivul este acela că respectiva variabilă de substituţie nu conţine nici un fel de informaţii noi care să nu fie incluse în celelalte variabile de substituţie.
- SPSS generează automat variabile de substituţie pentru variabila dependentă, dar trebuie să aibă informaţii referitoare la care dintre variabilele predictori sunt nominale (variabile categoriale).
- Aplicaţia SPSS generează tabele de clasificare, acestea indicând predicţia apartenenţei la o anumită categorie, pe baza variabilelor predictori. Acesta reprezintă un indicator de încredere al calităţii predicţiei, deoarece se precizează şi numărul de clasificări corecte.
- Analizele prin regresie logistică includ şi un număr mare de statistici de tip *goodness-of-fit*, bazate pe coeficientul chi-square. Acestea au mai multe funcţii, dar cea mai importantă este reprezentată de faptul că indică îmbunătăţirea corespondenţei dintre categoria de apartenenţă prezisă şi apartenenţa reală la o categorie. Un predictor util ar trebui să îmbunătăţească gradul de corespondenţă dintre apartenenţa prezisă şi categoriile reale cărora le aparţin diferitele cazuri.
- Regresia logistică multinomială trebuie folosită atunci când există trei sau mai multe categorii pentru variabila dependentă (prezisă sau criteriu). Dacă există doar *două* categorii pentru variabila dependentă, atunci se foloseşte, de obicei, regresia logistică binomială (vezi capitolul 33).

Utilizarea regresiei multinomiale poate fi ilustrată folosind informaţiile descrise în tabelul 32.1 (*ISP*, tabelul 36.2). Aceste date provin dintr-un studiu fictiv al diferenţelor

Tabelul 32.1. *Date pentru regresia logistică multinomială*

	Age	DAS	Mother hostile	Father hostile	Children's home	Physical abuse	Sexual abuse	Type of offence
1	younger	low	high	low	no	yes	no	rapist
2	younger	low	high	low	no	yes	yes	rapist
3	older	low	high	low	no	yes	yes	rapist
4	older	high	high	high	yes	no	no	incest
5	older	high	high	high	yes	yes	yes	rapist
6	younger	low	high	low	no	no	no	rapist
7	older	high	low	high	no	yes	yes	rapist
8	older	high	low	high	yes	no	no	incest
9	younger	low	low	high	yes	no	yes	incest
10	older	high	high	low	no	yes	yes	incest
11	older	high	low	low	yes	no	yes	incest
12	younger	high	low	high	no	yes	no	rapist
13	older	high	low	high	yes	no	yes	incest
14	older	high	high	low	yes	yes	yes	incest
15	older	low	high	high	no	yes	yes	incest
16	younger	high	high	low	yes	no	no	paedophile
17	older	high	low	high	yes	no	yes	paedophile
18	older	low	high	high	no	no	yes	paedophile
19	younger	high	low	high	yes	yes	yes	paedophile
20	older	low	low	high	yes	no	no	paedophile
etc.								

dintre violatori, infractori sexuali incestuoşi şi pedofili. Aceasta înseamnă că pentru categoriile de infractori avem o variabilă nominală sau categorială cu trei valori diferite. În acest exemplu, toate variabilele predictori - vârsta, scala DAS (scala depresiei, anxietăţii şi stresului), ostilitatea mamei, ostilitatea tatălui, căminul copiilor, abuzul fizic şi abuzul sexual - sunt variabilele nominale (categorială) care au doar două valori diferite în fiecare caz. Trebuie să subliniem faptul că orice tip de variabilă poate fi folosit ca predictor în regresia logistică multinomială. Totuşi, cercetătorul trebuie să indice care sunt variabilele numerice din analiză.

32.1. Introducerea datelor

Aceste date sunt introduse în SPSS în maniera obişnuită, fiecare coloană reprezentând o variabilă. În scopuri didactice, informaţiile au fost repetate de 10 ori, pentru a se ajunge la un set de date realist, dar şi pentru a limita, în acelaşi timp, munca persoanelor care doresc să reproducă analiza cu exactitate.

Introduceţi datele.

Ponderaţi cazurile după valoarea „freq".

	age	das	motherhs	fatherhs	childhom	physabus	sexualab	typoff	freq	var
1	1	1	2	1	2	1	2	1	10	
2	1	1	2	1	2	1	1	1	10	
3	2	1	2	1	2	1	1	1	10	
4	2	2	2	2	1	2	2	2	10	
5	2	2	2	2	1	1	1	1	10	
6	1	1	2	1	2	2	2	1	10	
7	2	2	1	2	2	1	1	1	10	
8	2	2	1	2	1	2	2	2	10	
9	1	1	1	2	1	2	1	2	10	
10	2	2	2	1	2	1	1	2	10	
11	2	2	1	1	1	2	1	2	10	
12	1	2	1	2	2	1	2	1	10	
13	2	2	1	2	1	2	1	2	10	
14	2	2	2	1	1	1	1	2	10	
15	2	1	2	2	2	1	1	2	10	
16	1	2	2	1	1	2	2	3	10	
17	2	2	1	2	1	2	1	3	10	
18	2	1	2	2	2	2	1	3	10	
19	1	2	1	2	1	1	1	3	10	
20	2	1	1	2	1	2	2	3	10	
21										

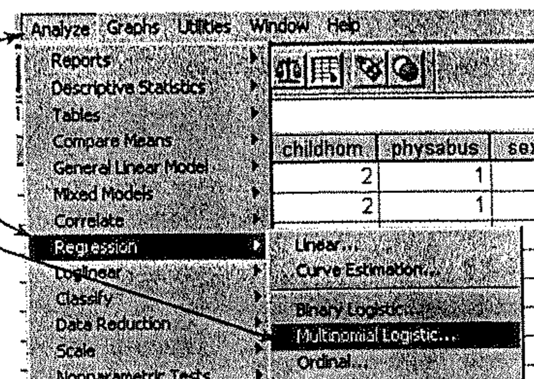
În cazul tuturor analizelor prin regresie logistică, variabilele dummy sunt create de SPSS. Acest lucru se întâmplă întotdeauna în cazul variabilelor prezise (în acest caz, tipul de infracţiune), dar se poate realiza şi în cazul variabilelor-predictori nominale (categoriale) care au mai mult de două categorii. Dacă situaţia o necesită, SPSS generează variabile noi (variabile de substituţie) care vor fi incluse în analiză. Aşadar, nu trebuie să vă surprindă faptul că veţi găsi în output variabile care nu făceau parte din input-ul iniţial de date. SPSS nu afişează în foaia de lucru variabilele dummy pe care le creează - totuşi, acestea sunt menţionate în output. SPSS creează automat variabile de substituţie adecvate pentru variabila dependentă (criteriu), pe baza numărului de valori (categorii) diferite ale acelei variabile. Variabilele predictori sau independente sunt şi ele codificate cu variabile de substituţie dacă sunt definite de către cercetător ca fiind variabile nominale (categoriale). Variabilele dummy sunt tratate pe larg în sumarul de la începutul acestui capitol, ca şi în textul statistic adiacent (*ISP*, capitolul 36).

Înainte de versiunea 12 a aplicaţiei SPSS, regresia logistică multinomială nu avea decât un număr de opţiuni relativ nesatisfăcător. Totuşi, începând cu versiunea 12, SPSS conţine şi o versiune stepwise în cadrul căreia variabilele sunt selectate ca predictori în ordinea puterii de predicţie independente pe care o au. În acest ghid, vom analiza datele folosind regresia logistică multinomială stepwise. Constatările diferă puţin la nivelul detaliilor de cele din textul statistic adiacent (*ISP*), dar diferenţele nu sunt substanţiale.

32.2. Regresia logistică multinomială stepwise

Pasul 1

Selectaţi „Analyze", „Regression" şi „Multinomial Logistic...".

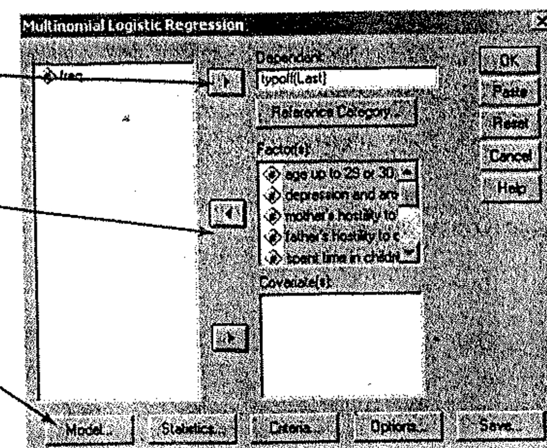


Pasul 2

Selectaţi „typoff" şi apăsaţi butonul ► de lângă caseta „Dependent:" pentru a introduce opţiunea în casetă.

Selectaţi celelalte şapte variabile fie separat, fie împreună (excluzând „freq") şi apăsaţi butonul ◄ de lângă caseta „Factor(s):" pentru a le introduce.

Selectaţi „Model...".

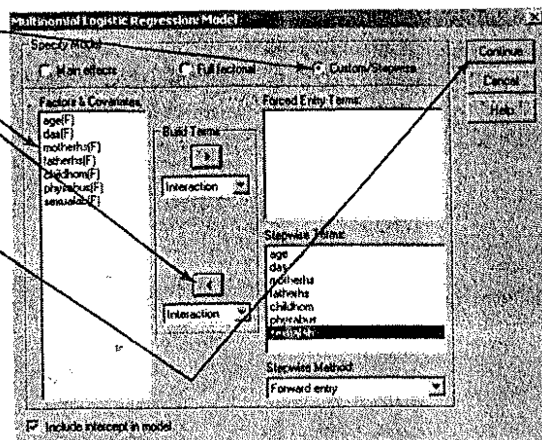


Pasul 3

Selectați „Custom/ Stepwise". Selectați fiecare dintre cele șapte variabile pe rând și apăsați butonul ◄ de lângă meniul „Stepwise terms:" pentru a le introduce acolo.

Selectați „Continue".

Selectați „Statistics..." din ecranul anterior, care va reapărea.

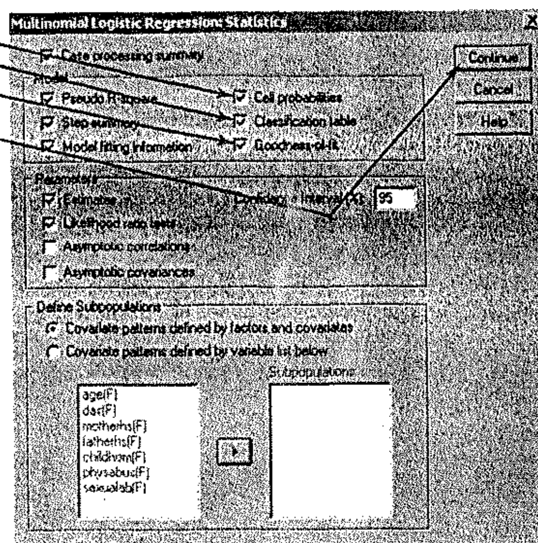


Pasul 4

Selectați „Cell probabilities", „Classification table" și „Goodness-of-fit".

Selectați „Continue".

Apăsați „OK" în ecranul anterior care va reapărea.



32.3. Interpretarea output-ului

Regresia nominală

Atenție

Există 37 (64,9%) celule (adică nivele de variabile dependente pe subpopulații) cu frecvența zero.
Procedura stepwise a fost întreruptă deoarece s-a descoperit o problemă numerică.
Este posibil să existe o separație aproape completă a datelor. Fie nu există date estimative referitoare la probabilitatea maximă, fie unele valori estimative ale parametrilor sunt infinite. Procedura NOMREG va continua în ciuda avertismentului(elor) de mai sus. Rezultatele afișate ulterior se bazează pe ultima iterație a ultimului model introdus în procedura stepwise. Valabilitatea corespondenței modelului nu este sigură.

Case Processing Summary

		N	Marginal Percentage
type of offence	rapist	70	35.0%
	Incestuous child abuser	80	40.0%
	paedophile	50	25.0%
age up to 29 or 30 and above	younger	70	35.0%
	older	130	65.0%
depression and anxiety scale	low score	80	40.0%
	high score	120	60.0%
mother's hostility to offender as child	low hostility	90	45.0%
	high hostility	110	55.0%
father's hostility to offender as child	how hostility	80	40.0%
	high hostility	120	60.0%
spent time in children's home	yes	110	55.0%
	no	60	45.0%
physically abused as a child	yes	100	50.0%
	No	100	50.0%
Sexually abused as a child	yes	130	65.0%
	no	70	35.0%
Valid		200	100.0%
Missing		0	
Total		200	
Subpopulation		19[a]	

a. Variabila dependentă are o singură valoare observată la 18 (94,7%) subpopulații.

Acest tabel prezintă distribuția fiecărei variabile din analiză. Distribuția infracțiunilor este încercuită.

Acest tabel arată ordinea în care au fost selectate variabilele-predictori.

Step Summary

Model	Action	Effect(s)	-2 Log Likelihood	Chi-Square	df	Sig.
0	Entered	Intercept	407.957			
1	Entered	childhom	330.132	77.825	2	.000
2	Entered	age	300.651	29.481	2	.000
3	Entered	physabus	274.195	26.456	2	.000
4	Entered	sexualab[a]	258.708	15.487	2	.000

Metoda stepwise: introducerea progresivă
a. Procedura stepwise a fost întreruptă deoarece a apărut o problemă de natură numerică.

Modelul din pasul 0 nu conține variabile-predictori. În pasul 1, se introduce variabila „childhom". Aceasta este cel mai bun predictor pe care îl avem la dispoziție.

„Sexualab" este ultima variabilă introdusă în ecuația de regresie. Analiza nu mai continuă după acest exemplu.

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	407.957			
Final	258.708	149.250	8	.000

Modelul „Intercept Only" nu conține *nici una* dintre variabilele predictori. Modelul „Final" este cel care folosește grupul format din cei mai buni predictori. Observați că există o îmbunătățire semnificativă a corespondenței dintre date atunci când se folosește modelul „Final".

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	240.587	28	.000
Deviance	255.236	28	.000

Acest tabel indică măsura în care datele prezise corespund datelor reale. Există o diferență semnificativă între cele două – așadar, predicția nu este perfectă.

Pseudo R-Square

Cox and Snell	.526
Nagelkerke	.594
McFadden	.345

Acest tabel oferă trei valori estimative ale „corelației multiple" dintre variabilele-predictori și apartenența la una dintre categoriile infractorilor sexuali. În acest caz, valorile sunt moderate (1,00 ar indica o clasificare perfectă), ceea ce confirmă faptul că predicția este imperfectă.

Acesta este un tabel foarte important, care indică măsura în care îndepărtarea fiecăreia dintre variabilele-predictori din predicție reduce în mod semnificativ corespondența predicției cu datele reale. După cum se poate vedea, în acest caz, îndepărtarea oricăruia dintre cei patru predictori afectează în mod contrar corespondența datelor prezise cu datele actuale.

Likelihood Ratio Tests

Effect	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	258.708[a]	.000	0	.
age	275.878	17.170	2	.000
childhom	287.336	28.628	2	.000
physabus	298.372	39.664	2	.000
sexualab	274.195	15.487	2	.000

Statistica chi-square reprezintă diferența dintre probabilitățile -2 log dintre modelul final și un model redus. Modelul redus se formează prin omiterea unui efect din modelul final. Ipoteza nulă este aceea că toți parametrii acelui efect sunt 0.

a. Acest model redus este echivalent cu modelul final, deoarece omiterea efectului nu intensifică gradul de libertate.

Acest tabel indică ponderile *b* pe care computerul le va folosi atunci când realizează predicțiile. Ponderile *b* sunt utile pentru raportare, dar cercetătorul nu are, de fapt, nevoie să le folosească în calcule.

Parameter Estimates

type of offence[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B) Lower Bound	Upper Bound
rapist	Intercept	.608	.660	.851	1	.356			
	[age=1]	.218	.559	.152	1	.697	1.243	.416	3.718
	[age=2]	0[b]	.	.	0	.	.	.	.
	[childhom=1]	-2.252	.545	17.042	1	.000	.105	.036	.306
	[childhom=2]	0[b]	.	.	0	.	.	.	.
	[physabus=1]	20.124[b]	.613	1077.1	1	.000	****	165110941.9	1826627708
	[physabus=2]	0	.	.	0	.	.	.	.
	[sexualab=1]	-18.582	.000	.	1	.	.000	.000	.000
	[sexualab=2]	0[b]	.	.	0	.	.	.	.
Incestuous child abuser	Intercept	.326	.565	.334	1	.563			
	[age=1]	-1.533	.455	11.330	1	.001	.216	.088	.527
	[age=2]	0[b]	.	.	0	.	.	.	.
	[childhom=1]	.122	.486	.063	1	.802	1.130	.436	2.929
	[childhom=2]	0[b]	.	.	0	.	.	.	.
	[physabus=1]	.782	.496	2.485	1	.115	2.185	.827	5.774
	[physabus=2]	0[b]	.	.	0	.	.	.	.
	[sexualab=1]	.299	.455	.431	1	.512	1.348	.552	3.291
	[sexualab=2]	0[b]	.	.	0	.	.	.	.

a. Categoria de referință este: pedofil.

b. Valoarea acestui parametru este stabilită la zero pentru că este un parametru redundant.

Atunci când interpretați output-ul, valorile variabilei trebuie verificate cu atenție. Ponderea *b* este -1,533 pentru categoria 1,00 a variabilei vârstă – celeilalte categorii îi este asociată o pondere de 0,00. Eticheta de valoare de 1,00 din acest caz înseamnă că infractorul era mai tânăr. Totuși, există și o valoare negativă a lui *b*. Acest lucru înseamnă că infractorii incestuoși care abuzează de copii se disting, în parte, de celelalte două grupuri, prin faptul că sunt mai în vârstă. Acest lucru poate crea confuzii la început, dar este pur și simplu modul în care operează aplicația SPSS.

Eroarea standard este de 0,00, ceea ce înseamnă că valoarea Wald ar fi infinit de mare. Nici unul dintre nivelele de semnificație nu poate fi procesat, deși această variabilă este foarte semnificativă din punct de vedere statistic.

Tabelul de clasificare indică într-o manieră simplă cât de bună a fost predicția. După cum se poate vedea, predictorii sunt foarte performanți în ceea ce privește predicția corectă a violatorilor (85,7% cazuri corecte). Predictorii nu prezic bine pedofilii (doar 20% dintre pedofili sunt identificați corect).

Classification

Observed	Predicted rapist	Incestuous child abuser	paedophile	Percent Correct
rapist	60	10	0	85.7%
Incestuous child abuser	20	50	10	62.5%
paedophile	10	30	10	20.0%
Overall Percentage	45.0%	45.0%	10.0%	60.0%

Aici sunt afişate toate combinaţiile posibile de valori ale variabilelor-predictori. Prin urmare, acest tabel poate fi folosit pentru identificarea celui mai probabil rezultat prezis, în cazul oricărei combinaţii de variabile-predictori. Adică modelul predictor al unui anumit infractor ar putea fi folosit pentru a identifica cea mai probabilă predicţie pentru modelul respectiv.

Observed and Predicted Frequencies

								Frequency			Percentage	
Sexually abused as a child	physically abused as a child	spent time in children's home	father's hostility to offender as child	mother's hostility to offender as child	depression and anxiety scale	age up to 29 or 30 and above	type of offence	Observed	Predicted	Pearson Residual	Observed	Predicted
yes	yes	yes	low hostility	high hostility	high score	older	rapist	0	1.367	-1.269	.0%	13.9%
							incestuous child abuser	10	7.078	2.032	100.0%	70.8%
							paedophile	0	1.535	-1.347	.0%	15.3%
			high hostility	low hostility	high score	younger	rapist	0	3.801	-2.372	.0%	38.0%
							incestuous child abuser	0	3.192	-2.165	.0%	31.9%
							paedophile	10	3.207	4.603	100.0%	32.1%
				high hostility	high score	older	rapist	10	1.387	7.862	100.0%	13.9%
							incestuous child abuser	0	7.078	-4.922	.0%	70.8%
							paedophile	0	1.535	-1.347	.0%	15.3%
		no	low hostility	high hostility	low score	younger	rapist	10	8.601	1.328	100.0%	86.0%
							incestuous child abuser	0	.702	-.869	.0%	7.0%
							paedophile	0	.797	-.930	.0%	8.0%
						older	rapist	10	6.281	2.433	100.0%	62.8%
							incestuous child abuser	0	2.987	-2.084	.0%	29.9%
							paedophile	0	.732	-.889	.0%	7.3%
					high score	older	rapist	0	6.281	-4.110	.0%	62.8%
							incestuous child abuser	10	2.987	4.845	100.0%	29.9%
							paedophile	0	.732	-.889	.0%	7.3%
			high hostility	low hostility	high score	older	rapist	10	6.281	2.433	100.0%	62.8%
							incestuous child abuser	0	2.987	-2.084	.0%	29.9%
							paedophile	0	.732	-.889	.0%	7.3%
				high hostility	low score	older	rapist	0	6.281	-4.110	.0%	62.8%
							incestuous child abuser	10	2.987	4.845	100.0%	29.9%
							paedophile	0	.732	-.889	.0%	7.3%
	No	yes	low hostility	low hostility	high score	older	rapist	0	.000	.000	.0%	.0%
							incestuous child abuser	10	6.785	2.177	100.0%	67.9%
							paedophile	0	3.215	-2.177	.0%	32.1%
			high hostility	low hostility	low score	younger	rapist	0	.000	.000	.0%	.0%
							incestuous child abuser	10	3.130	4.686	100.0%	31.3%
							paedophile	0	6.870	-4.686	.0%	68.7%
					high score	older	rapist	0	.000	.000	.0%	.0%
							incestuous child abuser	10	13.571	-1.710	50.0%	67.9%
							paedophile	10	6.429	1.710	50.0%	32.1%
		no	high hostility	high hostility	low score	older	rapist	0	.000	.000	.0%	.0%
							incestuous child abuser	0	6.514	-4.323	.0%	65.1%
							paedophile	10	3.486	4.323	100.0%	34.9%
no	yes	no	low hostility	high hostility	low score	younger	rapist	10	10.000	.000	100.0%	100.0%
							incestuous child abuser	0	.000	.000	.0%	.0%
							paedophile	0	.000	.000	.0%	.0%
			high hostility	low hostility	high score	younger	rapist	10	10.000	.000	100.0%	100.0%
							incestuous child abuser	0	.000	.000	.0%	.0%
							paedophile	0	.000	.000	.0%	.0%
	No	yes	low hostility	high hostility	high score	younger	rapist	0	1.623	-1.340	.0%	16.2%
							incestuous child abuser	0	2.141	-1.651	.0%	21.4%
							paedophile	10	6.336	2.403	100.0%	63.4%
			high hostility	low hostility	low score	older	rapist	0	.701	-.868	.0%	7.0%
							incestuous child abuser	0	5.675	-3.622	.0%	56.7%
							paedophile	10	3.624	4.194	100.0%	36.2%
					high score	older	rapist	0	.701	-.868	.0%	7.0%
							incestuous child abuser	10	5.675	2.761	100.0%	56.7%
							paedophile	0	3.624	-2.364	.0%	36.2%
				high hostility	high score	older	rapist	0	.701	-.868	.0%	7.0%
							incestuous child abuser	10	5.675	2.761	100.0%	56.7%
							paedophile	0	3.624	-2.364	.0%	36.2%
		no	low hostility	high hostility	low score	younger	rapist	10	8.374	2.385	100.0%	83.7%
							incestuous child abuser	0	.835	-.954	.0%	8.3%
							paedophile	0	2.791	-1.967	.0%	27.9%

Procentele se bazează pe totalul frecvenţelor observate în cazul fiecărei subpopulaţii.

În cazul regresiei logistice multinomiale, output-ul este unul substanţial. Desigur, este posibil să reducem numărul de rezultate, dar în acest caz trebuie să specificaţi cu claritate care dintre aspectele output-ului nu sunt esenţiale obiectivului dumneavoastră. Din moment ce este mai uşor să ignorăm tabelele care sunt în plus decât să refacem analiza, este de preferat să greşim în plus şi să avem un număr prea mare de rezultate decât invers.

- Este bine să vă amintiţi să examinaţi tabelul intitulat „Case Processing Summary". Acesta vă va reaminti distribuţia categoriilor fiecărei variabile din analiză. În exemplul nostru, toate variabilele au doar două categorii, şi de aceea nu este necesară crearea variabilelor de substituţie. Totuşi, este posibil ca variabilele predictori să aibă trei sau mai multe categorii, caz în care SPSS va crea automat variabile de substituţie corespunzătoare pentru toate categoriile respectivei variabile, în afară de una. (Această metodă nu se aplică în cazul variabilelor care sunt definite drept covariate, acestea fiind tratate ca variabile numerice.) Există totuşi trei valori diferite ale variabilei criteriu a tipului infracţiunii – violatori, infractori incestuoşi şi pedofili, iar aplicaţia SPSS va crea două variabile de substituţie pentru aceste trei categorii.

- Tabelul „Step Summary" indică, practic, secvenţa variabilelor introduse în regresia multiplă stepwise. Amintiţi-vă că în cazul analizei stepwise, predictorii sunt selectaţi în funcţie de puterea lor (diferită) de predicţie. Aşadar, cel mai bun predictor este selectat primul, după ce sunt făcute ajustările necesare, apoi se alege cel mai bun predictor dintre cei rămaşi şi aşa mai departe. Există şi un pas 0, care conţine doar intercepţia liniei de regresie. În acest exemplu, paşii 1-4 aduc pe rând cei mai puternici predictori. Variabila „childhom" este adăugată în pasul 1, iar „vârsta" – în pasul 2. „Physabus" este adăugat în pasul 3, iar „sexualab" – în pasul 4. Fiecare dintre aceşti paşi determină o corespondenţă semnificativ mai mare între datele prezise (modelate) şi cele reale. Acest lucru se poate vedea prin micşorarea valorilor probabilităţilor -2 log (care sunt valori chi-square) date în tabel. Fiecare dintre aceste modificări este semnificativă în acest exemplu, ceea ce înseamnă că nici unul dintre predictori nu poate fi eliminat fără a periclita acurateţea clasificării.
- Tabelul „Model Fitting Information" dă valoarea probabilităţii -2 log a coeficientului chi-square pentru corespondenţa modelului (adică predictorii semnificativi plus intercepţia). Această valoare este semnificativă în acest caz. Nu este decât un semn că modelul (variabilele predictori) nu prezice în totalitate datele reale. Cu alte cuvinte, predicţia este incompletă sau parţială. În mod evident, există şi alţi factori care trebuie luaţi în considerare pentru obţinerea unei corespondenţe perfecte între model şi date. De obicei, aceasta este situaţia.
- Tabelul statisticii „Pseudo R^2" nu face decât să confirme acest lucru. Statistica R^2 oferă corelaţia combinată dintre un set de predictori şi variabilele prezise pentru datele numerice. Coeficientul pseudo R^2 este similar acestui procedeu în ceea ce priveşte interpretarea, dar este folosit atunci când nu este posibilă procesarea corectă a statisticii R^2 propriu-zise, aşa cum se întâmplă în cazul regresiei logistice. După cum se poate vedea, s-au folosit trei metode diferite de calcul. Toate cele trei sunt determinate de relaţia combinată dintre predictori şi variabilele categoriale. Valoarea zero înseamnă că nu există corelaţie multiplă, valoarea 1,00 semnifică o corelaţie multiplă perfectă. Valorile situate în jurul numărului 0,5 sunt destul de satisfăcătoare, deoarece indică o corelaţie combinată generală dintre variabilele predictori şi variabila prezisă cu o valoare de aproximativ 0,7. (Aceasta se obţine prin extragerea rădăcinii pătrate a valorii 0,5.)
- Tabelul „Likelihood Ratio Tests" ne relatează ce se întâmplă cu modelul dacă eliminăm, pe rând, fiecare dintre variabilele-predictori. Modelul nu este altceva decât un set de predictori care apar în analiză. În acest caz, avem patru predictori, aşa cum s-a specificat deja. În fiecare caz, există o descreştere semnificativă la nivelul corespondenţei dintre datele prezise şi cele reale, în urma eliminării unuia dintre predictori. Cu alte cuvinte, fiecare dintre predictori are un efect semnificativ şi ar

trebui, în mod normal, să fie păstrat. Desigur, dacă cercetătorul are motive întemeiate, oricare predictor poate fi eliminat, deşi cercetătorilor neexperimentaţi li se recomandă să nu facă acest lucru.

- Tabelul „Parameter Estimates" indică, practic, ponderile de intercepţie şi regresie în cazul acestei analize de regresie multinomială. Valoarea intercepţiei este de exact 0,608 în cazul violatorilor. Dar observaţi o serie de lucruri. Variabila dependentă (tipul de infracţiune), care are trei categorii, a fost transformată în două variabile de substituţie „violatori" (faţă de celelalte două grupuri) şi „agresori incestuoşi care molestează copii" (faţă de celelalte două grupuri). Ţineţi minte că numărul de variabile de substituţie se obţine din numărul de categorii minus unu. Există trei categorii de infractori, aşadar vom avea două variabile de substituţie. Variabilele dummy sunt create prin luarea uneia dintre cele trei categorii de infractori şi compararea cu restul categoriilor de infractori. În exemplul nostru, avem:

  - violatori faţă de agresori incestuoşi care molestează copii *şi* pedofili;
  - agresori incestuoşi care molestează copii faţă de violatori *şi* pedofili;
  - pedofili faţă de violatori *şi* agresori incestuoşi care molestează copii.

  Alegerea uneia dintre cele trei variabile de substituţie posibile este una arbitrară şi poate fi variată prin selectarea opţiunii „Custom/Stepwise" în pasul 3 al secţiunii 32.2. De asemenea, observaţi că variabilelor li s-au asociat două ponderi de regresie – câte una pentru fiecare valoare. Totuşi, una dintre perechi este întotdeauna zero, ceea ce, practic, înseamnă că aceste valori nu au contribuit cu nimic la realizarea calculului. Semnificaţia fiecăreia dintre aceste ponderi de regresie este dată de una dintre coloanele din tabel. Semnificaţia se bazează pe valoarea Wald care este dată într-o altă coloană. Nu trebuie să vă faceţi probleme dacă nu înţelegeţi foarte bine acest lucru din moment ce nu este nevoie să faceţi calcule propriu-zise. Metoda este explicată în detaliu în textul statistic adiacent (*ISP*, capitolul 36).
- Tabelul „Classification" este foarte important şi dă precizie predicţiilor, pe baza parametrilor estimativi. Acest tabel de asociere indică ce predicţii ar putea fi realizate pe baza variabilelor-predictori semnificative şi cât de precise ar putea fi aceste predicţii. După cum se poate vedea, predicţiile sunt foarte precise în cazul violatorilor şi mai puţin exacte în cazul agresorilor incestuoşi care molestează copii. Totuşi, clasificarea este slabă în cazul pedofililor.
- Tabelul „Observed and Predicted Frequencies" este, probabil, folositor mai ales pentru cei care au o situaţie practică în care doresc să facă cea mai bună predicţie a categoriei pe baza variabilelor predictori. Tabelul oferă toate modelele posibile de variabile predictori (în SPSS acestea se numesc covariate) şi clasificările reale ale datelor pentru fiecare model şi afişează şi cel mai probabil final prezis pe baza respectivului model. Cu alte cuvinte, tabelul ar putea fi folosit pentru a realiza predicţii pentru cazurile individuale, pe baza modelelor individuale cunoscute.

32.4. Raportarea constatărilor

Nu există nici o metodă convenţională de raportare succintă a constatărilor în cazul regresiei logistice multinomiale. Dacă tehnica a mai fost folosită în studii anterioare din respectiva arie de cercetare, puteţi adopta aceleaşi metode de raportare utilizate anterior.

În cazul analizei de date realizate în acest capitol, este evident că există o serie de predictori care funcţionează în parte relativ eficient. O modalitate de raportare a acestor constatări ar fi următoarea: „Regresia logistică multinomială a fost efectuată pentru a stabili care sunt caracteristicile care disting cele trei tipuri de infractori. S-a demonstrat că patru dintre cele şapte variabile-predictori incluse în analiză diferenţiază într-o anumită măsură cele patru tipuri distincte de infractori. Violatorii au fost identificaţi corect ca atare în 85,7% din cazuri, agresorii incestuoşi care molestează copii au fost identificaţi corect în 62,5% din exemple, dar pedofilii au fost identificaţi corect doar în 20% din situaţii. S-a observat o tendinţă foarte frecventă de a categorisi greşit pedofilii ca fiind infractori incestuoşi care molestează copiii. Predictorii care au diferenţiat cel mai bine violatorii au fost timpul petrecut într-o casă unde locuiesc copii (b = –2,252), abuzul fizic (b = 20,124) şi abuzul sexual (b = –18, 582). Ultimul predictor nu este reportat în tabel ca fiind semnificativ ca atare. Nu i se acordă nici un fel de semnificaţie. Totuşi, pedofilii au fost cel mai bine diferenţiaţi de celelalte două grupe doar după vârstă (b = –1,533)".

33 Regresia logistică binomială

Rezumat

- Regresia logistică binomială poate fi considerată a fi un caz particular al regresiei logistice multinomiale care a fost descrisă în capitolul 32. Distincția esențială este aceea că diferențiază caracteristicile persoanelor în doar *două* grupuri diferite.
- Aplicația SPSS are un repertoriu mult mai bogat de tehnici de regresie pentru regresia logistică binomială, deși versiunile 12 și 13 au adus puțin mai multă flexibilitate procedurilor referitoare la regresia logistică multinomială.
- Regresia logistică binomială este o formă de regresie multiplă. Aceasta identifică modele de variabile care pot diferenția cu eficacitate membrii a două categorii diferite. Așadar, regresia logistică binomială prezice apartenența la o anumită categorie la fel cum o face o valoare numerică în cazul regresiei multivariate (capitolul 28). De exemplu, puteți folosi această metodă pentru examinarea modelului de variabile care face diferența cel mai bine între bărbații și femeile care participă la un studiu asupra motivelor de consultare a medicului. Poate apărea un model al motivelor pentru care bărbații merg la doctor diferit de modelul motivelor pentru care femeile fac acest lucru.
- Altfel spus, regresia logistică binomială folosește variabile predictori pentru a prezice căreia dintre categoriile variabilei dependente este mai probabil să aparțină diferiți indivizi.
- Procedura de regresie logistică binomială calculează valorile *b* (sau ponderile de regresie), în mod asemănător metodei folosite în regresia multiplă (capitolul 28). Marea diferență este aceea că în cazul regresiei logistice binomiale aceste ponderi *b* nu se aplică pentru prezicerea scorurilor. În schimb, se aplică în cazul unei mărimi numite *logit*, care este logaritmul natural al raportului șanselor. Raportul șanselor este asemenea unei probabilități. Nu este altceva decât raportul dintre numerele dintr-o categorie și numărul cazurilor din altă categorie. În general, valorile *logit* (ca, de altfel, și logaritmii naturali) nu sunt esențiale în utilizarea regresiei logistice binomiale de către un cercetător, și de aceea nu este necesară înțelegerea profundă a acestui concept.
- Deși categoriile sunt întotdeauna variabile categoriale binomiale (binară sau cu două valori), celelalte variabile din analiză pot fi variabile numerice, nominale (categoriale) sau pot fi o combinație între cele două tipuri.
- Orice variabilă categorială care are mai mult de două categorii este transformată în mod automat de către SPSS într-un grup de variabile de substituție. Din motive tehnice, una dintre variabilele dummy posibile este omisă în mod arbitrar din regresia logistică binomială. Acest lucru se întâmplă deoarece nu conține nici un fel de informații noi care să nu fie deja prezente în celelalte variabile de substituție.

- Regresia binomială, de exemplu, poate începe prin folosirea procedurilor stepwise, introducerea progresivă a predictorilor slabi, eliminarea regresivă a predictorilor slabi și așa mai departe. Alegerea tipului de model care trebuie folosit este de multe ori o chestiune destul de subtilă. Procedurile stepwise construiesc gradat modelul prin selectarea pas cu pas a variabilelor care pot să diferențieze bine între membrii celor două categorii ale variabilei dependente. Introducerea progresivă înseamnă că, odată cu fiecare pas, computerul identifică variabila rămasă care reușește cel mai bine să diferențieze între cele două categorii. Dacă nu sunt îndeplinite anumite condiții în ceea ce privește puterea de predicție (sau clasificare), atunci respectiva variabilă nu este introdusă în model. În psihologie, metoda introducerii progresive stepwise este o alegere bună pentru realizarea multor studii.
- O mare parte din output-ul SPSS al regresiei logistice binomiale constă în factori care indică măsura în care datelor modelate (apartenența la o anumită categorie prezisă de către variabilele-predictori) corespund datele reale (categoria reală căreia îi aparține individul). Acești indicatori se bazează în mare măsură pe coeficientul chi-square. Un predictor util ar trebui să îmbunătățească corespondența dintre apartenența prezisă și categoriile reale cărora le aparțin cazurile.
- Mai ușor de înțeles sunt tabelele de clasificare generate de SPSS, care indică măsura în care predicțiile corespund clasificărilor reale.

Exemplul pe care îl vom folosi este un studiu al variabilelor care ar putea fi folosite la evaluarea probabilității ca un prizonier să comită o nouă infracțiune sau nu după ieșirea din închisoare. Comiterea unei noi infracțiuni este cunoscută sub numele de recidivă. Datele din studiu sunt prezentate în tabelul 33.1 (*ISP*, tabelul 37.5). Din motive pedagogice și pentru comoditatea celor care doresc să urmeze pe computer pașii pe care-i efectuăm, cele 19 cazuri sunt reproduse de cinci ori. După cum se poate vedea, recidiva este o variabilă categorială binomială – într-o anumită perioadă de timp, prizonierul fie recidivează, fie nu. Din moment ce scopul analizei noastre este acela de a identifica modele de variabile care prezic în care dintre aceste două categorii se va încadra un prizonier, atunci acesta constituie în mod evident un set de date pentru regresia logistică binomială.

Orice tip de variabilă – numerică sau nominală (categorială) – poate fi folosit ca variabilă-predictor în regresia logistică binomială. Totuși, în exemplul nostru, am folosit variabile foarte simple, cum ar fi vârsta, detenția anterioară, tratamentul, remușcările, starea civilă și faptul că infractorul a comis sau nu o infracțiune sexuală. Toate acestea sunt variabile nominale binomiale. Putem folosi variabile numerice, care sunt definite drept covariate în SPSS. Putem folosi, de asemenea, variabile nominale (categoriale) având trei sau mai multe valori (categorii) diferite. În acest caz, SPSS recodifică variabila transformând-o automat într-o serie de variabile de substituție. Aceasta înseamnă că setul de variabile-predictori este, în realitate, mai mare decât numărul inițial de variabile din date. Dacă am avea o variabilă cum este tipul de infracțiune, aceasta având la rândul ei mai multe categorii, cum ar fi infracțiunile sexuale, infracțiunile violente și furtul, atunci, teoretic, am putea crea trei variabile de substituție. Acestea ar fi :

- Variabila dummy 1 = infracțiune sexuală sau nu ;
- Variabila dummy 2 = infracțiune violentă sau nu ;
- Variabila dummy 3 = furt sau nu.

În realitate, crearea variabilelor de substituţie este la fel de simplă cum este şi crearea variabilelor dummy. Există o singură complicaţie pe care trebuie să o avem în vedere. Una dintre variabilele de substituţie nu conţine, de fapt, nici o informaţie nouă dacă ştim care sunt valorile unei persoane referitoare la celelalte două variabile de substituţie.

Tabelul 33.1. *Date pentru studiul recidivei – datele preluate din 19 cazuri sunt reproduse de cinci ori, pentru a oferi un eşantion de dimensiuni realiste pentru simplificarea explicaţiilor*

	Recidivism	Age	Previous prison	Treatment	Contrite	Married	Sex offender
1	yes	younger	yes	no	no	no	yes
2	yes	older	yes	no	no	no	yes
3	yes	older	yes	yes	no	no	yes
4	yes	older	yes	yes	no	yes	no
5	yes	younger	yes	no	no	no	no
6	yes	younger	no	yes	yes	no	no
7	yes	older	no	yes	yes	yes	yes
8	yes	younger	yes	no	no	no	yes
9	yes	younger	no	no	no	yes	yes
10	yes	older	no	no	no	no	no
11	no	younger	no	yes	yes	no	no
12	no	older	no	yes	yes	no	no
13	no	older	yes	yes	yes	yes	yes
14	no	younger	no	yes	yes	yes	yes
15	no	younger	no	yes	yes	no	yes
16	no	younger	no	no	yes	yes	no
17	no	older	no	no	no	yes	no
18	no	older	yes	yes	yes	no	no
19	no	older	yes	yes	yes	no	no
etc.	yes	younger	yes	no	no	no	yes

Aşadar, atunci când ştim că o persoană nu a comis o infracţiune sexuală şi ştim că nu a comis nici o infracţiune violentă, înseamnă că respectiva persoană a comis un furt. Din moment ce *toţi* indivizii sunt infractori, acesta nu este decât un raţionament logic simplu. Din cauza faptului că întotdeauna există o variabilă de substituţie care se repetă şi este redundantă, în regresia logistică avem întotdeauna un număr cu unu mai mic decât maximum de variabile de substituţie. SPSS alege în mod arbitrar variabila dummy pe care să o elimine dacă utilizatorul nu face nici o precizare în acest sens.

Tabelul 33.2 ilustrează o modalitate de codificare a datelor pregătite pentru analiza cu ajutorul computerului. Cel mai bine este să introduceţi în SPSS datele sub formă numerică, astfel încât fiecărei variabile să i se asocieze un cod numeric. Cifrele 1 şi 0 au fost folosite în fiecare caz pentru a indica prezenţa sau absenţa unei caracteristici. Acestor valori li se vor da ulterior etichete valorice în SPSS, astfel încât output-ul să fie cât mai clar.

Regresia logistică binomială din SPSS are o varietate largă de opţiuni de analiză. Începătorilor li se va părea puţin derutantă mulţimea posibilităţilor şi multe dintre opţiuni nu vor prezenta nici un interes pentru mulţi cercetători. Alegerea opţiunilor adecvate (ceea ce unii ar numi modelul adecvat) depinde foarte mult de scopul analizei care se efectuează, iar cercetătorul este singurul care cunoaşte acest aspect. În general, dacă unicul scop al analizei este obţinerea celui mai bun set de predictori pentru clasificarea persoanelor, atunci este suficient să introduceţi toţi predictorii în acelaşi timp – SPSS ia pur şi simplu variabilele în ordinea în care sunt enumerate. Totuşi, de obicei, cercetătorul încearcă să construiască o explicaţie conceptuală/teoretică; de exemplu, vor să afle motivul pentru care unii infractori recidivează în acest exemplu. În aceste condiţii, este posibil ca cercetătorul să dorească să acorde prioritate în analiză unora dintre variabile. Aceasta implică introducerea ierarhică a predictorilor în blocuri. În plus,

Tabelul 33.2. *Datele din tabelul 33.1 codificate binar cu 0 şi 1 pentru fiecare variabilă*

	Recidivism	Age	Previous prison	Treatment	Contrite	Married	Sex offender
1	1	0	1	0	0	0	1
2	1	1	1	0	0	0	1
3	1	1	1	1	0	0	1
4	1	1	1	1	0	1	0
5	1	0	1	0	0	0	0
6	1	0	0	1	1	0	0
7	1	1	0	1	1	1	1
8	1	0	1	0	0	0	1
9	1	0	0	0	0	1	1
10	1	1	0	0	0	0	0
11	0	0	0	1	1	0	0
12	0	1	0	1	1	0	0
13	0	1	1	1	1	1	1
14	0	0	0	1	1	1	1
15	0	0	0	1	1	0	1
16	0	0	0	0	1	1	0
17	0	1	0	0	0	1	0
18	0	1	1	1	1	0	0
19	0	1	1	1	1	0	0
etc.	1	0	1	0	0	0	1

este necesară identificarea celor mai buni predictori. Variabilele pot fi introduse pas cu pas, cel mai bun predictor fiind luat primul, apoi cel mai bun dintre predictorii rămaşi este introdus al doilea (permiţând realizarea corelaţiilor dintre predictori) şi aşa mai departe, până când nu se mai poate realiza nici o îmbunătăţire semnificativă în ceea ce priveşte predicţia.

O altă posibilitate ar fi introducerea iniţială a tuturor predictorilor, apoi eliminarea lor unul câte unul. Dacă eliminarea celui mai slab predictor reduce semnificativ acurateţea

predicției, înseamnă că acesta trebuie păstrat. Totuşi, el poate fi eliminat *dacă* îndepărtarea sa nu produce nici o schimbare în exactitatea predicției. Procesul continuă pas cu pas, prin analizarea efectelor îndepărtării celui mai slab predictor dintre cei rămaşi. Există o singură complicație pe care trebuie să o aveți în vedere – aceea că predictorii eliminați pot, de fapt, să revină în rândul grupului de predictori dacă eliminarea unei alte variabile are ca rezultate creşterea puterii predictive independente a variabilei eliminate anterior.

În cele din urmă, eliminarea regresivă *nu* este mai bună decât alte abordări, ci, în mare, adoptă doar o metodă diferită. Nici nu va da neapărat exact aceleaşi rezultate ca alte metode – pur şi simplu, este o abordare rezonabilă şi adecvată a datelor.

33.1. Regresia logistică binomială

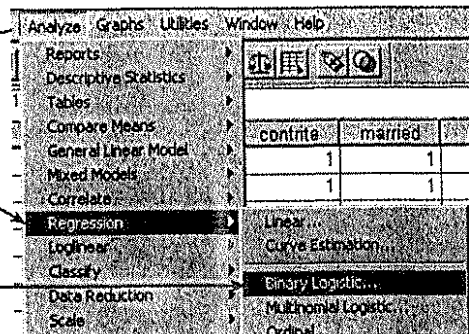
Pasul 1

Introduceți datele. Ponderați cazurile folosind variabila „freq".

	recidivi	age	prepris	treatmen	contrite	married	sexoff	freq
1	1	1	1	1	1	1	0	5
2	1	0	1	1	1	1	0	5
3	1	0	1	0	1	1	0	5
4	1	0	1	0	1	0	1	5
5	1	1	1	1	1	1	1	5
6	1	1	0	0	0	1	1	6
7	1	0	0	0	0	0	1	5
8	1	1	1	1	1	1	0	5
9	1	1	0	1	1	0	0	6
10	0	0	0	1	1	1	0	5
11	0	1	0	0	0	1	0	5
12	0	0	0	0	0	1	1	5
13	0	0	1	0	0	0	1	5
14	0	1	0	0	0	0	1	5
15	0	1	0	0	0	1	0	5
16	0	1	0	1	0	0	0	5
17	0	0	0	1	1	0	0	5
18	0	0	0	0	0	1	1	5
19	0	0	1	0	0	1	1	5

Pasul 2

Selectați „Analyze", „Regression" şi „Binary Logistic...".

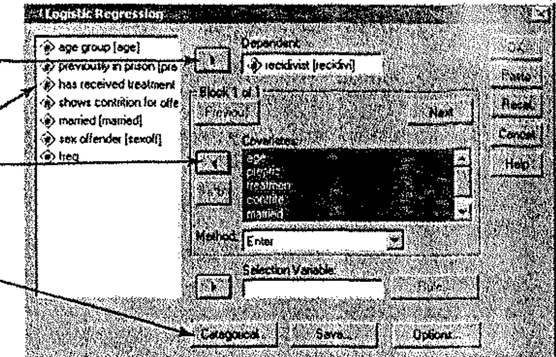


Pasul 3

Selectați „recidivi" şi apăsați butonul ▶ de lângă caseta „Dependent:" pentru a introduce opțiunea acolo.

Selectați celelalte şase variabile (în afară de „freq") şi apăsați butonul ◀ de lângă caseta „Covariates:" pentru a le introduce acolo.

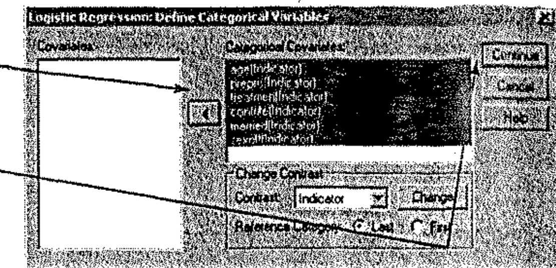
Selectați „Categorical...".



Pasul 4

Selectați toate cele şase variabile şi apăsați butonul ◀ de lângă caseta „Categorical Covariates:" pentru a le introduce acolo.

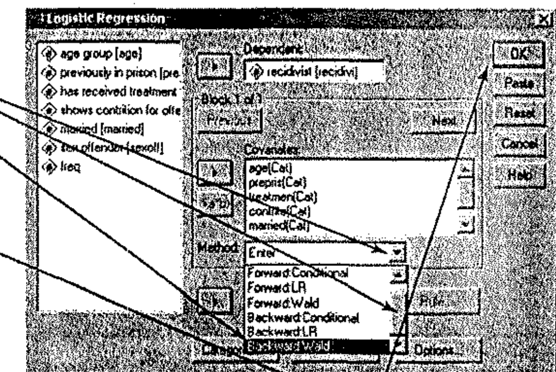
Selectați „Continue".



Pasul 5

Apăsați butonul ▼ din caseta „Method:", deplasați bara de derulare până vedeți opțiunea „Backward Wald" şi selectați-o.

Apăsați „OK".



33.2. Interpretarea output-ului

Regresia logistică

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	19	100.0
	Missing Cases	0	.0
	Total	19	100.0
Unselected Cases		0	.0
Total		19	100.0

a. Dacă ponderea este activată, consultați tabelul de clasificare pentru a verifica numărul total de cazuri.

Dependent Variable Encoding

Original Value	Internal Value
No	0
yes	1

Verificați acest tabel – explică modul exact în care a fost codificată variabila dependentă de către aplicația SPSS („Internal Value"). În acest caz, valorile interne sunt identice cu valorile pe care le-am introdus, dar merită să verificați, pentru prevenirea unor erori grave.

Categorical Variables Codings

		Frequency	Paramete (1)
sex offender	sex offender	10	1.000
	not sex offender	9	.000
previously in prison	no previous	11	1.000
	previous prison	8	.000
has received treatment in prison	treated	11	1.000
	not treated	8	.000
shows contrition for offence	Contrite	10	1.000
	not contrite	9	.000
married	married	7	1.000
	not married	12	.000
age group	older (30 plus)	10	1.000
	younger	9	.000

În acest tabel se descompune fiecare variabilă, specificându-se numărul exact de persoane din fiecare categorie. Așadar, există nouă persoane în eșantionul celor care nu sunt infractori sexuali.

Blocul 0: blocul inițial

Blocul 0 reprezintă momentul anterior luării în considerare a vreunuia dintre predictori. În general, puteți sări peste această etapă.

Classification Table[a,b]

Observed			Predicted: recidivist No	Predicted: recidivist yes	Percentage Correct
Step 0	recidivist	No	50	0	100.0
		yes	45	0	.0
	Overall Percentage				52.6

a. Constanta este inclusă în model.
b. Valoarea de rotunjire este de 0,500.

Acest tabel specifică numărul real de recidiviști și nerecidiviști. Există 45 de recidiviști și 50 de nerecidiviști.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-.105	.205	.263	1	.608	.900

Constanta este mai mult sau mai puțin aceeași ca și intercepția din cazul regresiei multivariate.

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	age(1)	2.299	1	.129
		prepris(1)	21.159	1	.000
		treatmen(1)	6.345	1	.012
		contrite(1)	31.714	1	.000
		married(1)	.452	1	.501
		sexoff(1)	.293	1	.588
	Overall Statistics		50.804	6	.000

În această etapă nu au fost incluși predictori, astfel încât nu poate apărea nici unul în ecuația de regresie.

Blocul 1: Metodă = stepwise regresiv (Wald)

Blocul 1 conține tabelele-cheie.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	70.953	6	.000
	Block	70.953	6	.000
	Model	70.953	6	.000
Step 2[a]	Step	-3.825	1	.050
	Block	67.128	5	.000
	Model	67.128	5	.000
Step 3[a]	Step	-17.310	1	.000
	Block	49.818	4	.000
	Model	49.818	4	.000
Step 4[a]	Step	-.116	1	.733
	Block	49.702	3	.000
	Model	49.702	3	.000
Step 5	Step	16.968	1	.000
	Block	66.670	4	.000
	Model	66.670	4	.000

a. O valoare negativă a lui chi-square indică faptul că valoarea chi-square a scăzut față de etapa anterioară.

Pasul 1 include toți predictorii.

Pasul 2 și pașii ulteriori de dedesubt implică eliminarea unei variabile – adică reprezintă pașii de eliminare regresivă.

Tabelul „Variables in the Equation" (care urmează) indică predictorii prezenți în model în fiecare etapă.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	60.482[a]	.526	.702
2	64.307[a]	.507	.676
3	81.617[b]	.408	.545
4	81.733[b]	.407	.544
5	64.765[a]	.504	.673

a. Estimarea s-a încheiat la iterația numărul 20, deoarece s-a atins numărul maxim de iterații.
Soluția finală nu a fost găsită.

b. Estimarea s-a încheiat la iterația numărul 6, deoarece parametrii estimativi s-au modificat cu mai puțin de 0,001.

- Aceștia sunt indicii de corespondență a modelului cu datele reale. Cu cât sunt mai apropiate de 1,0 cele două valori (pseudo) R^2, cu atât mai mare este corespondența modelului cu datele reale în etapa respectivă. Cu cât este mai mică valoarea probabilității –2 log, cu atât se potrivește mai bine modelul cu datele reale.

Observați că modelul final (pasul 5) are o corespondență mai mare – s-a permis reintrarea unei variabile în ecuația de regresie.

Classification Table[a]

	Observed		Predicted: recidivist No	Predicted: recidivist yes	Percentage Correct
Step 1	recidivist	No	45	5	90.0
		yes	5	40	88.9
	Overall Percentage				89.5
Step 2	recidivist	No	45	5	90.0
		yes	5	40	88.9
	Overall Percentage				89.5
Step 3	recidivist	No	50	0	100.0
		yes	10	35	77.8
	Overall Percentage				89.5
Step 4	recidivist	No	50	0	100.0
		yes	10	35	77.8
	Overall Percentage				89.5
Step 5	recidivist	No	45	5	90.0
		yes	5	40	88.9
	Overall Percentage				89.5

a. Valoarea de rotunjire este de 0,500.

Acest tabel indică acuratețea clasificării în fiecare etapă. Exactitatea generală rămâne constantă pe toată perioada – se schimbă doar proporțiile relative dintre recidiviștii și nerecidiviștii identificați corect. În funcție de obiectivele cercetătorului, este posibil să fie mai util unul dintre modelele anterioare. De exemplu, dacă este mai important să îi identificăm corect pe cei care nu vor recidiva, atunci modelele din pașii 3 și 4 sunt de preferat, deoarece acestea identifică nerecidiviștii cu acuratețe mai mare.

Cel mai puţin semnificativ predictor din pasul 1 se alege pentru eliminare în pasul 2. Atât „treatmen", cât şi „sexoff" au semnificaţii de 0,998, dar „treatmen" apare primul şi de aceea este ales.

Variables in the Equation[c]

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	age(1)	–2.726	.736	13.702	1	.000	.065
	prepris(1)	–1.086	.730	2.215	1	.137	.337
	treatmen(1)	19.362	8901.292	.000	1	.998	2.6E+08
	contrite(1)	–41.459	11325.914	.000	1	.997	.000
	married(1)	–.307	.674	.208	1	.648	.735
	sexoff(1)	–20.641	7003.093	.000	1	.998	.000
	Constant	23.802	7003.093	.000	1	.997	2.2E+10
Step 2[a]	age(1)	–2.775	.787	12.417	1	.000	.062
	prepris(1)	–1.471	.717	4.210	1	.040	.230
	contrite(1)	–22.879	7449.317	.000	1	.998	.000
	married(1)	–.448	.667	.450	1	.502	.639
	sexoff(1)	–21.102	7449.317	.000	1	.998	.000
	Constant	24.926	7449.317	.000	1	.997	6.7E+10
Step 3[a]	age(1)	–1.919	.710	7.306	1	.007	.147
	prepris(1)	–2.115	.686	9.496	1	.002	.121
	contrite(1)	–2.510	.617	16.539	1	.000	.081
	married(1)	.191	.560	.116	1	.733	1.210
	Constant	3.465	.940	13.600	1	.000	31.989
Step 4[a]	age(1)	–1.913	.711	7.247	1	.007	.148
	prepris(1)	–2.089	.685	9.299	1	.002	.124
	contrite(1)	–2.517	.616	16.687	1	.000	.081
	Constant	3.535	.926	14.566	1	.000	34.285
Step 5[b]	age(1)	–2.731	.786	12.077	1	.001	.065
	prepris(1)	–1.571	.695	5.107	1	.024	.208
	contrite(1)	–22.765	7466.473	.000	1	.998	.000
	sexoff(1)	–20.951	7466.473	.000	1	.998	.000
	Constant	24.648	7466.473	.000	1	.997	5.1E+10

a. Variabilă(e) introdusă(e) în pasul 1: „age", „prepris", „treatmen", „contrite", „married", „sexoff".
b. Variabilă(e) introdusă(e) în pasul 5: „sexoff".
c. Procedura stepwise a fost întreruptă deoarece îndepărtarea variabilei celei mai puţin semnificative determină producerea unui model stabilit anterior.

Deşi „sexoff" a fost eliminată din model începând cu pasul 3 şi nu apare în pasul 4, i se permite să reintre în model odată cu pasul 5 din simplul motiv că eliminarea variabilei „married" după pasul 3 a determinat creşterea puterii predictive independente a variabilei „sexoff" astfel încât să îndeplinească criteriile necesare pentru acceptarea din nou ca fiind un bun predictor.

Este important de observat că ponderile de regresie se aplică doar valorilor variabilei de categorie care este codificată cu 1. Aşadar, ponderea negativă de regresie pentru „sexoff(1)" de –20,951 indică, de fapt, că infractorii sexuali au o probabilitate *mai mică* de recidivă la eliberare. Trebuie să aveţi mare grijă să ştiţi ce codificări sunt asociate diferitelor valori ale variabilelor. Această problemă este agravată de faptul că variabilele nominale care nu sunt codificate cu 0 şi 1 pot fi recodificate de către SPSS, dându-li-se aceste valori.

Variables not in the Equation

			Score	df	Sig.
Step 2[a]	Variables	treatmen(1)	2.638	1	.104
	Overall Statistics		2.638	1	.104
Step 3[b]	Variables	treatmen(1)	6.160	1	.013
		sexoff(1)	13.026	1	.000
	Overall Statistics		14.017	2	.001
Step 4[c]	Variables	treatmen(1)	5.736	1	.017
		married(1)	.116	1	.733
		sexoff(1)	12.609	1	.000
	Overall Statistics		14.123	3	.003
Step 5[c]	Variables	treatmen(1)	2.809	1	.094
		married(1)	.453	1	.501
	Overall Statistics		2.994	2	.224

a. Variabilă(e) eliminată(e) în pasul 2: „treatmen".
b. Variabilă(e) eliminată(e) în pasul 3: „sexoff".
c. Variabilă(e) eliminată(e) în pasul 4: „married".

Acest tabel indică semnificaţia variabilelor care nu sunt incluse în model. Vedeţi care este semnificaţia variabilei „sexoff" în pasul 4. În consecinţă, această variabilă este reinclusă în pasul 5.

33.3. Raportarea constatărilor

O modalitate de raportare a constatărilor acestei analize ar fi următoarea: „Folosind modelul Wald al eliminării regresive, am analizat caracteristicile ce diferenţiază prizonierii care recidivează după eliberare de cei care nu comit o nouă infracţiune. Modelul final de regresie indică faptul că infractorii mai tineri, cei care au mai fost în detenţie, cei care nu au remuşcări în legătură cu infracţiunile lor şi cei care nu sunt infractori sexuali manifestă o probabilitate mai mare de recidivă. Vârsta şi detenţia anterioară au constituit predictori semnificativi în proporţie de 5%. Coeficientul Cox şi Snell pseudo R^2 a fost de 0,50, indicând doar o corespondenţă moderată între model şi datele reale. Acest model a fost aproape la fel de exact în ceea ce priveşte recidiva (corect în proporţie de 88,9%) ca în cazul nonrecidivei (90,0%)".

ANEXA A

Intervalele de încredere

- Intervalele de încredere nu constituie nici pe departe un concept nou în statistică, dar au fost de curând susţinute ca fiind preferabile estimărilor punctuale folosite în mod obişnuit în majoritatea analizelor statistice. Media unui eşantion poate fi folosită pentru estimarea mediei populaţiei căreia îi aparţine eşantionul. Această valoare estimată a mediei populaţiei este una punctuală din moment ce este exprimată ca unică valoare, cum este 5,3. Totuşi, ştim că această valoare estimată care se bazează pe un eşantion nu este decât cea mai bună posibilă şi că este probabil ca media reală a populaţiei să fie diferită de valoarea estimată.
- În cazul intervalelor de încredere, media estimată se exprimă în intervale de medii în care este probabil să se încadreze media populaţiei. Aşadar, intervalul de încredere al unei medii poate fi între 4,2 şi 6,4.
- Intervalul de încredere variază în funcţie de nivelul de certitudine cerut de cercetător. Aşadar, pentru aceleaşi date, intervalul de încredere care are o probabilitate de 95% să includă media populaţiei este mai mic decât intervalul de încredere cu o probabilitate de 99% de a conţine media populaţiei.
- Mai devreme, în această carte, am folosit intervalele de încredere atunci când păreau a fi utile în mod deosebit. Una dintre dificultăţile abordării cu intervale de încredere este constituită de faptul că, teoretic, orice statistică are un interval de încredere. În ciuda acestui lucru, este greu să găsim metodele de calculare a intervalelor în literatura de specialitate sau în manualele de statistică. Acest lucru limitează aplicabilitatea abordării cu intervale de încredere. Mai concret, aplicaţia SPSS, asemenea altor programe pe computer, nu oferă decât un număr limitat de intervale de încredere.

Intervalele de încredere au fost prezentate în output în cazul următoarelor teste care au fost descrise în această carte:

- Testul de regresie *B*: vedeţi tabelul de la pagina 86;
- Testul *t* pentru eşantioane perechi: vedeţi tabelul de la pagina 105;
- Testul *t* pentru eşantioane independente: vedeţi tabelul de la pagina 110;
- ANOVA One-way pentru eşantioane independente: vedeţi tabelul de la pagina 53;
- ANOVA Two-way pentru eşantioane independente: vedeţi tabelele de la paginile 163 şi 164;
- Testele prin comparaţii multiple: vedeţi tabelul de la pagina 170;
- ANCOVA One-way: vedeţi tabelul de la pagina 175.

Intervalele de încredere se pot obţine cu uşurinţă şi în cazul următoarelor statistici:

- Testul *t* pentru compararea unui eşantion cu media populaţiei One-Sample *t* test;
- ANOVA One-way pentru eşantioane perechi;
- ANOVA 2-way mixt;
- Regresia – pentru predicţia scorurilor.

ANEXA B

Alte metode statistice folosite în SPSS

Alte metode statistice oferite în SPSS şi care nu au fost descrise în această carte sunt enumerate mai jos, în funcţie de opţiunile din meniul şi submeniul „Analyze” şi de opţiunile din casetele de dialog.

Analyze menu	Analyze submenu	Dialog box
Descriptive Statistics	Crosstabs ...	Lambda
		Uncertainty coefficient
		Gamma
		Somers' d
		Kendall's tau-b
		Kendall's tau-c
		Risk
		Eta
Correlate	Bivariate ...	Kendall's tau-b
Regression	Curve Estimation ...	
	Binary Logistic ...	
	Multinomial Logistic ...	
	Ordinal ...	
	Probit ...	
	Nonlinear ...	
	Weight Estimation ...	
	2-Stage Least Squares ...	
	Optimal Scaling ...	
Loglinear	Logit ...	
Classify	TwoStep Cluster ...	
	K-Means Cluster ...	
	Hierarchical Cluster ...	
	Discriminant ...	
Data Reduction	Correspondence Analysis ...	
	Optimal Scaling ...	
Scale	Multidimensional Scaling ...	
	Multidimensional Scaling[PROXSCAL] ...	
	Multidimensional Scaling [ALSCAL]	
Nonparametric Tests	Binomial ...	
	Runs ...	
	1-Sample K-S ... (Kolmogorov-Smirnov)	

Analyze menu	Analyze submenu	Dialog box
	2 Independent Samples . . .	Kolmogorov–Smirnov Z
		Wald–Wolfowitz runs
		Moses extreme reactions
	K Independent Samples . . .	Kruskal–Wallis H
		Jonckheere–Terpstra
		Median
	2 Related Samples . . .	Marginal Homogeneity
	K Related Samples . . .	Friedman
		Kendall's W
		Cochran's Q
Time Series	Exponential Smoothing . . .	
	Autoregression . . .	
	ARIMA . . .	
	Seasonal Decomposition . . .	
Survival	Life Tables . . .	
	Kaplan–Meier . . .	
	Cox Regression . . .	
	Cox w/ Time-Dep Cov . . .	

Index

D

E

F

G

H

I

K

L

M

N

COLLEGIUM. Psihologie

au apărut:

Adrian Neculau (coord.) – *Psihologie socială. Aspecte contemporane*
Andrei Cosmovici – *Psihologie generală*
W. Doise, J.-C. Deschamps, G. Mugny – *Psihologie socială experimentală*
Adrian Neculau, Gilles Ferréol (coord.) – *Minoritari, marginali, excluși*
Constantin Cucoș – *Minciună, contrafacere, simulare. O abordare psihopedagogică*
Mielu Zlate (coord.) – *Psihologia vieții cotidiene*
R.Y. Bourhis, J.-P. Leyens (coord.) – *Stereotipuri, discriminare și relații intergrupuri*
Serge Moscovici – *Psihologia socială sau mașina de fabricat zei*
Adrian Neculau (coord.) – *Câmpul universitar și actorii săi*
Adrian Neculau (coord.) – *Psihologia câmpului social. Reprezentările sociale*
J. Barus-Michel, F. Giust-Desprairies, Luc Ridel – *Crize. Abordare psihosocială clinică*
Andrei Cosmovici, Luminița Iacob (coord.) – *Psihologie școlară*
Șerban Ionescu – *Paisprezece abordări în psihopatologie*
Serge Moscovici (coord.) – *Psihologia socială a relațiilor cu celălalt*
Willem Doise, Gabriel Mugny – *Psihologie socială și dezvoltare cognitivă*
Ana Stoica-Constantin, Adrian Neculau (coord.) – *Psihosociologia rezolvării conflictului*
André Sirota – *Conduite perverse în grup*
Adrian Neculau, G. Ferréol (coord.) – *Psihosociologia schimbării*
Mircea Miclea – *Psihologie cognitivă. Modele teoretico-experimentale*
Mielu Zlate – *Psihologia mecanismelor cognitive*
Gilles Ferréol (coord.) – *Identitatea, cetățenia și legăturile sociale*
Mielu Zlate – *Introducere în psihologie*
Ion Dafinoiu – *Elemente de psihoterapie integrativă*
Eva Drozda-Senkowska – *Psihologia socială experimentală*
Cornel Havârneanu – *Cunoașterea psihologică a persoanei. Posibilități de utilizare a computerului în psihologia aplicată*
Steve Duck – *Psihologia relațiilor interumane. Teorie, concepte și noțiuni fundamentale*
Valentin Clocotici, Aurel Stan – *Statistică aplicată în psihologie*
Gilles Ferréol (coord.) – *Adolescenții și toxicomania*
Adrian Neculau (coord.) – *Analiza și intervenția în grupuri și organizații*
Pierre De Visscher, Adrian Neculau (coord.) – *Dinamica grupurilor. Texte de bază*
Mihaela Roco – *Creativitate și inteligență emoțională*
Jean Delacour – *Introducere în neuroștiințele cognitive*
Ion Dafinoiu – *Personalitatea. Metode de abordare clinică: observația și interviul*
Septimiu Chelcea – *Un secol de cercetări în psihosociologie*
Aurel Stan – *Testul psihologic. Evoluție, construcție, aplicații*
Dan Goglează – *Psihoterapia ca relație a schimbării individuale. Strategii, cazuri, soluții, comentarii*
Alex Mucchielli – *Arta de a influența. Analiza tehnicilor de manipulare*
Michel-Louis Rouquette – *Despre cunoașterea maselor. Eseu de psihologie politică*
Vincent Yzerbyt, Georges Schadron – *Cunoașterea și judecarea celuilalt*
Mihai Kramar – *Psihologia stilurilor de gândire și acțiune umană*
Jean-Claude Abric – *Psihologia comunicării. Teorii și metode*
Șerban Ionescu, Marie-Madeleine Jacquet, Claude Lhote – *Mecanismele de apărare. Teorie și aspecte clinice*

Ştefan Boncu - *Psihologia influenţei sociale*
Daniel Gaonac'h, Pascale Larigauderie - *Memorie şi funcţionare cognitivă*
Bruna Zani, Augusto Palmonari - *Manual de psihologia comunităţii*
Ion Dafinoiu, Jenő László Vargha - *Hipnoza clinică. Tehnici de inducţie. Strategii terapeutice*
Adrian Neculau (coord.) - *Manual de psihologie socială*
Mihaela Vlăsceanu - *Organizaţii şi comportament organizaţional*
Iolanda Mitrofan (coord.) - *Cursa cu obstacole a dezvoltării umane. Psihologie, psihodiagnoză, psihoterapie centrată pe copil şi familie*
Constantin Enăchescu - *Tratat de psihanaliză şi psihoterapie*
Jacques Montangero - *Vis şi cogniţie*
Constantin Enăchescu - *Tratat de psihosexologie*
Roy Schafer - *Interpretarea psihanalitică a testului Rorschach*
Patrick Rateau - *Metodele şi statisticile experimentale în ştiinţele umane*
Ivana Marková - *Dialogistica şi reprezentările sociale*
Zoltán Bogáthy (coord.) - *Manual de psihologia muncii şi organizaţională*
Dorina Sălăvăstru - *Psihologia educaţiei*
Mielu Zlate - *Leadership şi management*
Constantin Enăchescu - *Tratat de igienă mintală*
Mielu Zlate - *Tratat de psihologie organizaţional-managerială*
Ticu Constantin - *Evaluarea psihologică a personalului*
Ana Stoica-Constantin - *Conflictul interpersonal. Prevenire, rezolvare şi diminuarea efectelor*
Alex Mucchielli - *Arta de a comunica. Metode, forme şi psihologia situaţiilor de comunicare*
Nicolae Mitrofan, Laurenţiu Mitrofan - *Testarea psihologică. Inteligenţa şi aptitudinile*
Ion Dafinoiu, Jenő-László Vargha - *Psihoterapii scurte. Strategii, metode, tehnici*
Constantin Enăchescu - *Tratat de teoria cercetării ştiinţifice*
Vasile Cernat - *Psihologia stereotipurilor*
Marius Milcu - *Psihologia relaţiilor interpersonale*
Constantin Enăchescu - *Tratat de psihopatologie*
Romeo Zeno Creţu - *Evaluarea personalităţii*
Gerald Matthews, Ian J. Deary, Martha C. Whiteman - *Psihologia personalităţii. Trăsături, cauze, consecinţe*
Daniel David - *Psihologie clinică şi psihoterapie. Fundamente*
Mihai Curelaru - *Reprezentări sociale*
Daniel David - *Metodologia cercetării clinice. Fundamente*
Dennis Howitt, Duncan Cramer - *Introducere în SPSS pentru psihologie*

în pregătire:

Ana Muntean - *Psihologia dezvoltării umane*
Zoltán Bogáthy (coord.) - *Manual de tehnici şi metode în psihologia muncii şi organizaţională*
Mielu Zlate - *Tratat de psihologie organizaţional-managerială*, vol. al II-lea
Ticu Constantin - *Psihologia reclamei*

www.polirom.ro

Redactor: Antonio Patraş
Coperta: Radu Răileanu
Tehnoredactor: Luminiţa Modoranu

Bun de tipar: august 2006. Apărut: 2006
Editura Polirom, B-dul Carol I nr. 4 • P.O. Box 266
700506, Iaşi, Tel. & Fax: (0232) 21.41.00; (0232) 21.41.11;
(0232) 21.74.40 (difuzare); E-mail: office@polirom.ro
Bucureşti, B-dul I.C. Brătianu nr. 6, et. 7, ap. 33,
O.P. 37 • P.O. Box 1-728, 030174
Tel.: (021) 313.89.78; E-mail: office.bucuresti@polirom.ro

Tiparul executat la S.C. LUMINA TIPO s.r.l.
str. Luigi Galvani nr. 20 bis, sect. 2, Bucureşti
Tel./Fax: 211.32.60, 212.29.27, E-mail: office@luminatipo.com